



TESIS - TE142599

PREDIKSI INDEKS HARGA KONSUMEN HARIAN MENGUNAKAN SUPPORT VECTOR REGRESSION BERBASIS CLOUD COMPUTING

RR. INTAN ARI BUDIASTUTI
NRP 2215206705

DOSEN PEMBIMBING
Mochamad Hariadi, ST., M.Sc., Ph.D
Dr. Supeno Mardi Susiki Nugroho, ST., MT

PROGRAM MAGISTER
BIDANG KEAHLIAN TELEMATIKA-CIO
JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017



TESIS - TE142599

PREDIKSI INDEKS HARGA KONSUMEN HARIAN MENGUNAKAN SUPPORT VECTOR REGRESSION BERBASIS CLOUD COMPUTING

RR. INTAN ARI BUDIASTUTI
NRP 2215206705

DOSEN PEMBIMBING
Mochamad Hariadi, ST., M.Sc., Ph.D.
Dr. Supeno Mardi Susiki Nugroho, ST., MT.

PROGRAM MAGISTER
BIDANG KEAHLIAN TELEMATIKA-CIO
JURUSAN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2017

LEMBAR PENGESAHAN

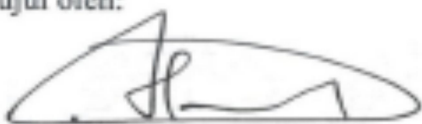
Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Teknik (M.T)
di
Institut Teknologi Sepuluh Nopember

Oleh:

Rr. Intan Ari Budiastuti
NRP. 2215206705

Tanggal Ujian : 6 Juni 2017
Periode Wisuda : September 2017

Disetujui oleh:

- 
1. Mochamad Hariadi, ST., M.Sc., Ph.D
NIP: 19691209 199703 1 002


(Pembimbing I)

- 
2. Dr. Supeno Mardi Susiki Nugroho, ST., MT
NIP: 19700313 199512 1 001


(Pembimbing II)

3. Dr. Ir. Endroyono, DEA
NIP: 19650404 199102 1 001


(Penguji)

- 
4. Dr. Eko Mulyanto Yuniarno, ST., MT.
NIP: 19680601 199512 1 009

(Penguji)

- 
5. Dr. Surya Sumpeno, ST., M.Sc.
NIP: 19690613 199702 1 003

(Penguji)

- 
6. Dr. Istas Pratomo, ST., MT.
NIP: 19790325 199702 1 003

(Penguji)

Dekan Fakultas Teknologi Elektro



Dr. Ir. Arief Sardjono, S.T., M.T.
NIP. 19760212 199512 1 001

Halaman ini sengaja dikosongkan

PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa isi keseluruhan Tesis saya dengan judul **“PREDIKSI INDEKS HARGA KONSUMEN HARIAN MENGGUNAKAN SUPPORT VECTOR REGRESSION BERBASIS CLOUD COMPUTING”** adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, Juni 2017

RR. INTAN ARI BUDIASTUTI
NRP. 2215206705

Halaman ini sengaja dikosongkan

PREDIKSI INDEKS HARGA KONSUMEN HARIAN MENGUNAKAN SUPPORT VECTOR REGRESSION BERBASIS CLOUD COMPUTING

Nama mahasiswa : Rr. Intan Ari Budiastuti
NRP : 2215206705
Pembimbing : 1. Mochamad Hariadi, ST., M.Sc., Ph.D
2. Dr. Supeno Mardi Susiki Nugroho, ST., MT

ABSTRAK

Inflasi yang parah bisa menyebabkan kelesuan perekonomian suatu negara. Oleh karena itu, inflasi perlu dikendalikan salah satunya dengan cara, setiap bulan Pemerintah melakukan prediksi dan perhitungan inflasi menggunakan indikator IHK. Prediksi dengan frekuensi bulanan, bisa jadi terlambat, karena inflasi sudah beberapa hari terjadi. Dengan berkembangnya teknologi internet saat ini maka berbagai sumber data berkaitan inflasi mudah diperoleh secara real-time dan ini dapat digunakan untuk data harian inflasi. Perkembangan varian dan volume data yang makin berkembang menimbulkan kebutuhan sistem komputasi yang baik. Cloud computing dapat digunakan untuk mengatasi permasalahan tersebut. Penelitian ini merupakan penelitian awal dalam mengembangkan model prediksi IHK harian menggunakan Big Data dan Cloud Computing. Didalam penelitian ini berfokus pada pengembangan prediksi IHK menggunakan data harian dengan metode Support Vector Regression (SVR) dan simulasinya di dalam Cloud Computing. Untuk pengembangan akurasi model yang lebih baik, dibandingkan fungsi kernel SVR dengan metode Linear Regression, Kernel Ridge Regression dan Random Forest dan melakukan optimasi parameter SVR menggunakan metode Grid Search. Hasil penelitian menunjukkan SVR dengan menggunakan kernel RBF mempunyai MSE lebih kecil yaitu sebesar 0.3454. Prediksi IHK harian ini disimulasikan ke dalam lingkungan cloud computing, hasil menunjukkan waktu proses di dalam cloud computing EC2 sebesar 0.1737 detik, lebih cepat dibandingkan personal computer sebesar 0.2746 detik.

Kata kunci: Support Vector Regression, Big Data, Cloud computing, Random Forest, Indeks Harga Konsumen

Halaman ini sengaja dikosongkan

PREDICTION OF DAILY CONSUMER PRICE INDEX USING SUPPORT VECTOR REGRESSION METHOD BASED CLOUD COMPUTING ENVIRONMENT

By : Rr. Intan Ari Budiastuti
Student Identity Number : 2215206705
Supervisor(s) : 1. Mochamad Hariadi, ST., M.Sc., Ph.D
2. Dr. Supeno Mardi Susiki Nugroho, ST., MT

ABSTRACT

Severe inflation can cause a country's economic downturn. Therefore, inflation needs to be controlled, one way, every month the Government predict and calculate inflation using CPI indicator. Prediction with monthly frequency, could be too late, because inflation has been a few days. With the development of internet technology today, various data sources related to inflation are easily obtained in real-time and this can be used for daily inflation data. The development of variants and the growing volume of data raises the need for a good computing system. Cloud computing can be used to solve the problem. This research is a preliminary research in developing daily CPI prediction model using Big Data and Cloud Computing. Here we focus on developing CPI predictions using daily data with the Support Vector Regression (SVR) method and its simulations in Cloud Computing. For better model accuracy development, compared to SVR kernel function with Linear Regression method, Kernel Ridge Regression and Random Forest and SVR parameter optimization using Grid Search method. The results showed that SVR using RBF kernel has a smaller MSE of 0.3454. Daily CPI predictions are simulated into cloud computing environments, the results show the processing time in cloud computing EC2 of 0.1737second faster than personal computer 0.2746 second..

Keywords: Support Vector Regression; Big Data; Cloud computing

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Alhamdulillah syukur kepada Allah SWT atas segala rahmat dan karunia yang telah dilimpahkan kepada penulis sehingga penulisan tesis dengan judul :

“PREDIKSI INDEKS HARGA KONSUMEN HARIAN MENGGUNAKAN SUPPORT VECTOR REGRESSION BERBASIS CLOUD COMPUTING”

dapat diselesaikan dengan baik. Buku tesis ini disusun untuk memenuhi salah satu syarat memperoleh gelar magister pada program studi Teknik Elektro dengan bidang keahlian Telematika CIO, Institut Teknologi Sepuluh Nopember.

Pada kesempatan ini penulis sampaikan terima kasih yang dalam kepada:

1. Kedua orangtua, suami, anak, yang selalu memberikan dorongan semangat dalam mengerjakan dan menyelesaikan tesis ini.
2. Prof.Dr.Ir. Budi Santosa M.Sc, Ph.D, Mochamad Hariadi, ST., M.Sc., Ph.D dan Dr. Supeno Mardi Susiki Nugroho, ST., MT atas bimbingan dalam menyelesaikan thesis ini.
3. Rekan-rekan seangkatan, Mahasiswa Program Studi Magister Telematika / Chief Information Officer (CIO) angkatan 2015 terima kasih atas kebaikan dan kerjasamanya dalam penelitian ini.
4. Rekan kerja Dinas Kominfo Pemkot Mojokerto bidang APIKA yang memberikan kesempatan saya melakukan study S2 dengan fokus

Penulis menyadari bahwa dalam penulisan tesis ini masih jauh dari kata sempurna, untuk itu demi perbaikan dan penyempurnaan tesis ini maka saran dan kritik membangun sangat diharapkan. Besar harapan penulis bahwa buku tesis ini dapat memberikan informasi dan manfaat bagi pembaca pada umumnya dan mahasiswa jurusan Teknik Elektro pada khususnya.

Surabaya, 6 Juni 2017

Penulis

Halaman ini sengaja dikosongkan

DAFTAR ISI

LEMBAR PENGESAHAN	iii
PERNYATAAN KEASLIAN TESIS	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xvii
DAFTAR TABEL	xix
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Batasan Masalah	5
1.5 Kontribusi Penelitian	5
1.6 Sistematika Penulisan	6
BAB 2 KAJIAN PUSTAKA	7
2.1 Prediksi di dalam Data Mining	7
2.1.1 Metode pemrosesan awal data atau Preprocessing	8
2.1.2 Cross-Validation	9
2.1.3 Metode Evaluasi	9
2.1.3.1 Explained variance score.	10
2.1.3.2 Mean absolute error	10
2.1.3.3 Median Absolute Error	10
2.2 Indeks Harga Konsumen	11
2.3 Support Vector Regression	13
2.4 Cloud Computing	19
2.4.1 Penyedia layanan cloud computing	22
2.4.2 Cara mengakses Cloud service	23
2.4.3 Migrasi aplikasi ke dalam Cloud Computing	25
2.5 Prediksi Finansial dengan menggunakan high frequency data	26
2.6 Linear Regression	27

2.7	Kernel Ridge Regression (KRR).....	28
2.8	Random Forest Regressor	29
2.9	Optimisasi Hyperparameter.....	32
2.9.1	Metode optimisasi hyperparameter Grid Search	33
2.9.2	Metode Optimisasi Hyperparameter Random Search	34
2.10	Tool komputasi mesin pembelajaran.....	35
BAB 3 METODOLOGI PENELITIAN		37
3.1	Tahapan Penelitian	37
3.2	Pemerolehan data mentah.....	38
3.3	Analisa data mentah yang akan digunakan dalam prediksi IHK harian..	41
3.4	Tahapan Pengolahan awal data atau Preprocessing	47
3.4.1	Mengubah nilai IHK ke dalam tahun dasar yang sama	47
3.4.2	Mengatasi permasalahan nilai yang hilang.....	47
3.4.3	Pengubahan skala data.....	48
3.5	Alokasi data.....	48
3.6	Pembangunan Model Prediksi IHK.....	49
3.6.1	Optimisasi Hyperparameter	49
3.6.2	Pembangunan prediksi IHK bulanan	51
3.6.3	Pembangunan prediksi harian.....	56
3.7	Tool penelitian.....	56
3.8	Skenario Simulasi Cloud Computing	57
3.8.1	Rancangan evaluasi prosedur dan fungsionalitas.	60
3.8.2	Rancangan evaluasi waktu komputasi metode optimisasi.....	61
3.8.3	Rancangan evaluasi waktu proses training dan testing	63
3.8.4	Rancangan evaluasi nilai akurasi prediksi IHK.....	64
BAB 4 HASIL DAN PEMBAHASAN		65
4.1	Analisa awal data.....	65
4.2	Hasil Preprocessing	69
4.3	Hasil pembangunan prediksi menggunakan data bulanan	69
4.4	Hasil pembangunan model prediksi IHK harian	81
4.5	Waktu komputasi dalam optimisasi hyperparameter	85
4.6	Hasil Evaluasi setelah migrasi prediksi ke Cloud	88
4.6.1	Hasil evaluasi prosedur dan fungsionalitas.....	88

4.6.2 Hasil evaluasi waktu komputasi optimisasi hyperparameter	90
4.6.3 Hasil Evaluasi Waktu Proses Training dan Testing data	92
4.6.4 Hasil evaluasi nilai MSE dan R^2	94
BAB 5 PENUTUP	95
5.1 Kesimpulan	95
5.2 Saran Penelitian Selanjutnya.....	96
DAFTAR PUSTAKA	97

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2.1 Cross-validation dengan $k=3$	9
Gambar 2.2 Arsitektur Support Vector Machine Regression	15
Gambar 2.3 Kernel mapping dari input space ke feature space	18
Gambar 2.4 Layanan Cloud beserta Penyediaanya.....	20
Gambar 2.5 Model Layanan Cloud.....	21
Gambar 2.6 Arsitektur Linear Regression	27
Gambar 2.7 Contoh Decision Tree.....	29
Gambar 2.8 Hasil dari Decision Tree regresi.....	32
Gambar 3.1 Diagram tahapan penelitian.....	37
Gambar 3.2 Cross Validation dengan 5-fold.....	49
Gambar 3.3 Diagram pembangunan prediksi IHK	52
Gambar 3.4 Proses estimasi IHK harian	54
Gambar 3.5. Flowchart pembangunan model menggunakan optimisasi Grid Search.....	55
Gambar 3.6 Tahapan migrasi prediksi IHK ke cloud computing	60
Gambar 4.2 Pengaruh hyperparameter epsilon dan degree terhadap nilai MSE... 75	75
Gambar 4.3 Grafik perbandingan rata-rata bulanan IHK harian menggunakan metode SVR kernel rbf dengan nilai aktual IHK	80
Gambar 4.4 Perbandingan Laju inflasi aktual dan hasil prediksi.....	81
Gambar 4.5 Perbandingan nilai MSE yang diperoleh saat testing prediksi IHK dengan lima metode prediksi	82
Gambar 4.6 Perbandingan waktu training prediksi IHK harian.....	83
Gambar 4.7 nilai IHK prediksi pada bulan Januari 2017.....	85
Gambar 4.8 Performansi metode Optimisasi hyperparameter	88
Gambar 4.9 Waktu komputasi metode Optimisasi	91
Gambar 4.10 Perbandingan waktu training 5 metode prediksi	93

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Fungsi Kernel.....	18
Tabel 2.2 Beberapa contoh Tipe Layanan Amazon.....	22
Tabel 2.3 Perbandingan Amazon Web Service dan Microsoft Azure	24
Tabel 2.4 Framework migrasi Cloud-RMM	25
Tabel 2.5 contoh data regresi	30
Tabel 2.6 Pembagian dataset berdasarkan atribut.....	31
Tabel 2.7 Standard deviasi dari data	31
Tabel 2.8 Kombinasi parameter menggunakan metode Grid Search.....	33
Tabel 3.1 Contoh Data Harga Konsumen Dari Siskaperbapo.com.....	39
Tabel 3.2 Contoh Data Harga Konsumen Dari Siskaperbapo.com (Lanjutan).....	40
Tabel 3.3 Ringkasan data nilai yang hilang di 8 kota Jawa Timur	42
Tabel 3.4 Ringkasan data nilai yang hilang di 8 kota Jawa Timur (Lanjutan)	43
Tabel 3.5 Ringkasan data yang hilang tiap variable	44
Tabel 3.6 Variable yang digunakan di dalam model prediksi IHK.....	45
Tabel 3.7 Contoh data bulanan Kota Surabaya dari Tahun 2012 - 2016	46
Tabel 3.8 Penerapan Grid Search pada Kernel Ridge Regression	51
Tabel 3.9 Perbedaan akun gratis AWS dan Azure.....	58
Tabel 3.10 Rancangan evaluasi prosedur dan fungsionalitas.....	61
Tabel 3.11 Rancangan evaluasi penggunaan metode Grid dan Random Search ..	62
Tabel 3.12 Rancangan evaluasi waktu proses training dan testing data	63
Tabel 3.13 Rancangan evaluasi nilai MSE dan R^2	64
Tabel 4.1 Nilai hilang dalam dataset harian kota Surabaya	65
Tabel 4.2 Dataset harian bulan November 2012 di Provinsi Jawa Timur	66
Tabel 4.3 Analisa harga telur ayam kampung di dataset Kota Malang	67
Tabel 4.4 Data kota Kediri tahun 2016 yang mengalami kekosongan nilai	68
Tabel 4.5 Hasil ringkasan nilai yang hilang pada 8 kota di Jawa Timur	69
Tabel 4.6. Perubahan IHK dengan basis tahun yang sama di Kota Surabaya	70
Tabel 4.7 Range hyperparameter yang digunakan.....	71
Tabel 4.8 Hasil R^2 dari Grid Search metode SVR-rbf.....	72
Tabel 4.9 Hasil optimisasi hyperparameter menggunakan Metode Grid Search..	73
Tabel 4.10. Hasil dari metode Random Search untuk SVR-RBF	74
Tabel 4.11 Hasil nilai mse setiap nilai hyperparameter C	76
Tabel 4.12 Hasil evaluasi pembangunan prediksi IHK bulanan	77
Tabel 4.13 Selisih antara nilai aktual IHK dengan rata bulanan IHK prediksi.....	78
Tabel 4.14 Performansi IHK hasil prediksi terhadap IHK aktual.....	79
Tabel 4.15 Perbandingan laju inflasi aktual dan prediksi	80
Tabel 4.16 Nilai MSE dan waktu komputasi prediksi IHK harian	82
Tabel 4.17 Data Hasil Prediksi Bulan Januari 2017	84
Tabel 4.18 Waktu komputasi dan best score dari metode optimisasi	86

Tabel 4.19 evaluasi prosedur dan fungsionalitas.....	89
Tabel 4.20 Hasil evaluasi penggunaan metode Grid dan Random Search.....	92
Tabel 4.21 Hasil evaluasi nilai MSE dan R^2	94

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Selama beberapa tahun terakhir, pemerintah telah mulai menggunakan teknologi berbasis web untuk mendukung aktifitas pemerintahan. Dalam hal ekonomi, pemerintah melakukan kegiatan ekonomi dan mempublikasikan kegiatannya di website resmi pemerintahan. Salah satu kegiatan yang berkaitan dengan ekonomi adalah prediksi dan perhitungan inflasi. Bank Indonesia mempublikasikan data-data berkaitan dengan ekonomi agar bisa diakses oleh masyarakat dan pihak yang mempunyai kepentingan. Inflasi juga menjadi perhatian Bank Indonesia. Karena inflasi yang parah bisa menyebabkan kelesuan perekonomian suatu negara, pemerintahan membuat terobosan untuk mengendalikan masalah inflasi dengan mengeluarkan kebijakan pengendalian inflasi di mulai dari daerah. Salah satu bentuk pengendalian inflasi yang dilakukan oleh pemerintah adalah setiap bulan melakukan prediksi inflasi dan perhitungan inflasi, menggunakan indikator IHK. Prediksi dengan frekuensi bulanan, bisa jadi terlambat, karena inflasi sudah beberapa hari terjadi dan tidak diketahui dengan cepat. Dengan berkembangnya teknologi internet saat ini maka berbagai sumber data berkaitan inflasi mudah diperoleh secara real-time. Data ini dapat dimanfaatkan untuk prediksi IHK harian. Prediksi IHK harian memungkinkan pengambil kebijakan membuat kebijakan lebih baik.

Di Indonesia, perhitungan inflasi dilakukan Badan Pusat Statistik, sedangkan prediksi inflasi dipublikasikan oleh Bank Indonesia. Setiap bulan, BPS mempublikasikan tingkat inflasi dan nilai Indeks Harga Konsumen. IHK merupakan salah satu indikator ekonomi yang menunjukkan rata-rata perubahan harga produk dan layanan yang dibayarkan oleh konsumen. Perubahan nilai IHK digunakan untuk mengukur tingkat inflasi yang terjadi. Untuk menghasilkan nilai IHK, BPS melakukan survey di beberapa daerah perwakilan di Indonesia. Sebelum tahun 2014 inflasi dihitung dari 66 kota di Indonesia sedangkan sejak tahun 2014 perhitungan inflasi menggunakan data 82 kota di Indonesia. Di Jawa

Timur diambil 8 kota untuk perhitungan statistik inflasi yaitu : Kota Surabaya, Kota Madiun, Kota Kediri, Kabupaten Sumenep, Kabupaten Jember, Kota Probolinggo, Kabupaten Banyuwangi dan Kota Malang. Sampai saat ini, BPS belum mempublikasikan nilai IHK harian, sehingga pembuat kebijakan di daerah tidak dapat mengetahui tingkat inflasi daerah mereka lebih cepat.

Perkembangan data digital, baik variable sumber ataupun volume yang semakin meningkat cepat, memungkinkan data inflasi tidak hanya diperoleh dari survey, tetapi bisa diperoleh dari sosial media dan media internet lainnya. Sebagai contoh, banyak hypermarket yang mempublikasikan harga barang yang dijual di website. Banyak juga masyarakat yang mendiskusikan permasalahan ekonomi di media sosial seperti Twitter dan Facebook. Twitter data dan harga komoditi di website hypermarket mungkin bisa digunakan sebagai variable perhitungan dan prediksi IHK. Teknik scrapping web dapat digunakan, untuk memperoleh data digital ini. Tidak perlu menunggu administrator website, menyediakan fitur data API untuk mengambil data tersebut. Dengan ini, prediksi inflasi IHK dapat dibangun dan dianalisa apakah ada hubungan antara data di media sosial yang mendiskusikan perekonomian dengan tingkat inflasi.

Teknologi yang bisa mengelola data yang terus bertumbuh ini dibutuhkan. Teknologi Big Data dapat digunakan untuk menghitung dan memprediksi data besar. Strategi penggunaan Big Data untuk perhitungan IHK telah dibahas oleh penelitian Doran [1]. Doran mengkaitkan permasalahan inflasi dengan Big Data, termasuk penggunaan teknik scrapping web.

Beberapa penelitian sebelumnya juga banyak membahas pengembangan prediksi inflasi [2], [3], [4], [5], [6]. Banyak variabel ataupun metode yang digunakan dalam penelitian tersebut. Dalam [2] Wulandari memperkirakan laju inflasi bulanan dengan membuat perbandingan model yang sesuai untuk memprediksi tingkat inflasi menggunakan metode ARIMA, varian Kalender dan Interferensi. Dalam [3], [4], [5], [6] menggunakan metode Artificial Neural Network (ANN) sebagai metode prediksi. Penelitian [3] membandingkan prediksi inflasi dengan menggunakan metode ARIMA dan ANN, dan menunjukkan bahwa, error prediksi dari ANN jauh lebih sedikit dibandingkan prediksi dengan menggunakan metode ARIMA. Dalam [6] Rita menggunakan ANN dan Genetika

algorithm (GA) dalam model prediksinya. Dijelaskan bagaimana membuat prediksi dari beberapa variabel yang mempengaruhi inflasi untuk meramalkan tingkat inflasi tahunan di Indonesia. Sedangkan di dalam penelitian di Cina [5] Linyun Zhang telah membandingkan tiga metode prediksi inflasi; Support Vector Machine untuk Regresi (SVR), BackPropagasi dan Linear Regression. Hasil dari penelitian tersebut menyimpulkan bahwa SVR memiliki prediksi yang lebih baik daripada BackPropagasi dan Linear Regression. Dalam penelitian tersebut belum membandingkan performansi antara kernel SVR. Penelitian inflasi diatas belum ada yang membahas prediksi menggunakan data harian.

Untuk kasus prediksi yang berkaitan dengan data yang besar, bisa merujuk penelitian Navin [7] yang membandingkan penggunaan Big Data analitik metode Support Vector Regression dan Decission Tree Regressor untuk peramalan harga emas. Dan penelitian dari Robin yang mengembangkan metode Random Forest (merupakan turunan dari decission tree) untuk permasalahan Big Data [8]. Dari penelitian penelitian sebelumnya dapat disimpulkan bahwa Support Vector Regression merupakan metode yang baik dan bisa digunakan untuk prediksi inflasi IHK dan metode Random Forest bisa digunakan sebagai metode pembanding didalam memproses Big Data.

Di dalam pengolahan data yang besar, sumber daya komputasi yang handal sangat dibutuhkan. Pemerintah sering menghadapi permasalahan biaya. Pembangunan data center untuk sektor admnistrasi publik, perawatan, keamanan IT dan sumber daya IT. Biaya implementasi yang dibutuhkan mungkin sangat besar sehingga bisa menyebabkan proyek IT tersebut gagal dilaksanakan. Untuk mengatasi permasalahan tersebut, Cloud computing mulai banyak diimplementasikan di dalam pemerintahan. [9]. Cloud computing menawarkan scalability yaitu kemudahan dalam memilih spesifikasi harware dan performansi hardware bisa ditingkatkan sesuai keinginan pengguna, penghematan biaya dengan penggunaan cloud dan keuntungan lainnya. Oleh karena itu penggunaan cloud computing memungkinkan Pemerintahan memiliki Data center sendiri tanpa kuatir permasalahan perangkat keras, legalisasi perangkat lunak, administrator IT dan lainnya.

Big Data mewakili konten dan cloud computing mewakili ketersediaan infrastruktur. Munculnya Big Data dalam cloud computing ditinjau detil dalam penelitian Ibrahim yang membahas hubungan antara data besar dan cloud computing dimana data besar digabungkan dengan cloud computing untuk memproses data secara terdistribusi dan cloud computing menyediakan engine ini. [10]

Di dalam penelitian ini, prediksi IHK dibangun sebagai solusi yang diajukan dalam menghadapi permasalahan diatas. Dengan pembangunan prediksi IHK harian, permasalahan yang telah dijelaskan diatas, akan kebutuhan IHK harian dan kendala kesiapan komputasi, dapat diatasi. Penelitian ini fokus pada pembangunan prediksi IHK harian menggunakan metode regresi Support Vector. Sebagai tambahan, metode Linear Regression dan Kernel Ridge Regression digunakan sebagai pembanding dengan alasan metode Linear regression sering digunakan dalam kasus regresi dan Kernel Ridge Regression merupakan pengembangan metode Linear yang memungkinkan dapat mengatasi permasalahan multicollinearity (keterikatan yang kuat antara variable prediksi). Dan satu lagi metode pembanding yang digunakan adalah Random Forest untuk regresi. Kemudian dilakukan eksperimen, membandingkan prediksi tersebut pada saat dijalankan di lingkungan cloud computing dan di lingkungan personal komputer

1.2 Rumusan Masalah

Permasalahan yang diangkat pada penelitian ini adalah :

Prediksi inflasi saat ini masih dilakukan dalam waktu bulanan padahal prediksi bulanan mungkin bisa terlambat. Selain itu, data yang terkait dengan inflasi masih belum terintegrasi sehingga untuk proses prediksi masih berjalan sendiri atau menunggu pihak lain.

1.3 Tujuan

Tujuan yang akan dicapai pada penelitian ini adalah membangun prediksi IHK menggunakan data harian yaitu harga bahan pokok harian dan kurs rupiah

terhadap dollar menggunakan Support Vector Regression. Kemudian mengintegrasikan semua sumber data ke dalam lingkungan cloud computing untuk mempercepat pemrosesan data.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini yaitu :

1. Penelitian ini hanya menggunakan data IHK sumber BPS dari Januari 2012 – Desember 2016
2. Data harian yang digunakan adalah data harga pokok dari 8 kota, bersumber dari website Sistem Informasi Harga Bahan Pokok Dinas Perindustrian dan Perdagangan Provinsi Jawa Timur. Penelitian ini hanya menggunakan harga bahan pokok tahun 2012 – 2016. Data 8 kota dianalisa dan hanya menggunakan satu data kota Surabaya untuk membangun prediksi IHK dan simulasi prediksi IHK di dalam Cloud Computing. Sebagai tambahan, variable nilai tukar mata Uang Rupiah terhadap dollar dijadikan data harian
3. Penelitian ini melakukan prediksi menggunakan data online dan belum membahas melakukan prediksi secara real time.
4. Dalam penelitian ini simulasi cloud computing dibatasi hanya pada migrasi prediksi IHK harian dan menghitung perbandingan waktu komputasi berdasarkan besar data, metode prediksi dan metode pencarian parameter prediksi.
5. Jenis cloud provider yang akan digunakan yaitu Amazon EC2 dan Microsoft Azure
6. Sistem Operasi yang akan digunakan yaitu Ubuntu 14.04
7. Penelitian yang dilakukan tidak membahas dari sisi keamanan sistem

1.5 Kontribusi Penelitian

Kontribusi dari penelitian ini yaitu :

Memberikan referensi bagi instansi terkait, menggunakan data harian untuk memprediksi IHK. Hasil penelitian ini diharapkan bisa sebagai bahan pertimbangan memprediksi inflasi dengan lebih baik dan mendapatkan gambaran

implementasi aplikasi e-government jika diimplementasikan ke dalam lingkungan cloud computing.

1.6 Sistematika Penulisan

Bab II

Bab ini berisi tentang landasan teori yang digunakan membangun mesin rekomendasi dan menjawab permasalahan penelitian.

Bab III

Bab ini berisi tentang metodologi dan metode yang digunakan untuk membangun dan menguji prediksi IHK harian dan menjelaskan skenario simulasi cloud computing

Bab IV

Bab ini berisi tentang hasil yang didapatkan dari prediksi IHK harian yang dibangun di personal computer dan cloud computing serta evaluasi dan analisis terhadap hasil dan evaluasi yang didapatkan.

Bab V

Bab ini berisi kesimpulan dan saran mengenai hasil evaluasi dan analisis serta penelitian selanjutnya yang dapat memperbaiki hasil penelitian ini.

BAB 2

KAJIAN PUSTAKA

2.1 Prediksi di dalam Data Mining

Data mining sering juga disebut *knowledge discovery in Database* (KDD) menurut buku Data Mining karya Budi santosa [14] adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini dapat digunakan untuk memperbaiki pengambilan keputusan di masa depan. *Machine learning* adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik yang bisa diprogramkan dan belajar dari data masa lalu. *Machine learning* menjadi alat analisis dalam data mining.

Metode pembelajaran dalam data mining secara garis besar dibagi menjadi dua pendekatan yaitu *supervised* dan *unsupervised*. Dalam pendekatan pertama, *unsupervised learning*, metode pembelajaran yang diterapkan tanpa adanya proses latihan (training) dan tanpa adanya target atau guru dalam hal ini label dari data. Contoh metode ini adalah *klastering* dan *Self Organizing Map* (SOM). Pendekatan yang kedua adalah *Supervised learning* yaitu metode belajar dengan adanya pelatihan. Metode yang termasuk dalam pendekatan kedua ini adalah Artificial Neural Network (ANN), Support Vector Machine (SVM). Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan data yang mempunyai label atau target untuk dilatih. Setelah dilatih dengan cukup data, maka mesin pembelajaran, misalkan SVM, bisa digunakan untuk meramal atau memprediksi keputusan apa yang diambil bila data yang baru masuk dan belum diketahui keluarannya. Jika performansi pada saat testing belum memenuhi harapan, nilai parameter *machine learning* bisa diatur untuk mendapatkan model dengan performansi yang lebih baik. Dalam hal SVM, parameter yang biasanya diatur adalah parameter C, epsilon, gamma.

Tugas peramalan atau prediksi di dalam data mining dapat dikategorikan ke dalam dua kelompok besar yaitu klasifikasi dan pendekatan fungsi atau dikenal

dengan regresi. Dalam klasifikasi, keluaran dari setiap data adalah bilangan bulat atau diskrit, sedangkan dalam regresi keluaran dari setiap data adalah bilangan kontinu. Perbedaan tugas klasifikasi dan regresi perlu ditegaskan karena mempunyai konsekuensi yang sangat penting, yaitu pemodelan matematika yang berbeda. Di dalam penelitian ini menggunakan pendekatan fungsi, yaitu prediksi mencari nilai kontinu.

Tahapan di dalam menerapkan data mining, mengadopsi SEMMA (dikembangkan oleh SAS Institute) kependekan dari Sample, Explore, Modify, Model dan Assess. Tahapan data mining terdiri dari :

1. Pengambilan data, data yang digunakan harus menampung informasi penting. Jika data mempunyai target kelas maka data yang diambil harus mewakili kelas tersebut secara berimbang
2. Analisa data, mengeksplorasi data
3. Modifikasi, atau preprocessing, sebelum menggunakan data dengan metode yang dipilih, sebaiknya dilakukan preprocessing terhadap data. Hal ini dilakukan untuk mendapatkan hasil yang akurat.
4. Model, memodelkan data dengan menggunakan alat analisis untuk mencari kombinasi data yang dapat memprediksi data uji dengan handal
5. Assess atau Evaluasi, mengevaluasi model prediksi yang telah dibangun.

2.1.1 Metode pemrosesan awal data atau Preprocessing

Pemrosesan awal data atau lebih dikenal dengan *preprocessing* ini perlu dilakukan untuk mendapatkan hasil analisis yang lebih akurat dalam penggunaan teknik machine learning. *Preprocessing* bisa dilakukan dengan cara melakukan transformasi nilai data, membuat nilai data menjadi lebih kecil skalanya tanpa merubah informasi didalamnya. Ada beberapa cara transformasi data, yang dilakukan dalam penelitian ini adalah *scalling*. *Scalling* adalah merubah data sehingga berada dalam skala tertentu. Skala ini berada diantara batas bawah dan batas atas, misalkan (0,1). Jika nilai maksimum tiap kolom adalah $\max(x_i)$ dan

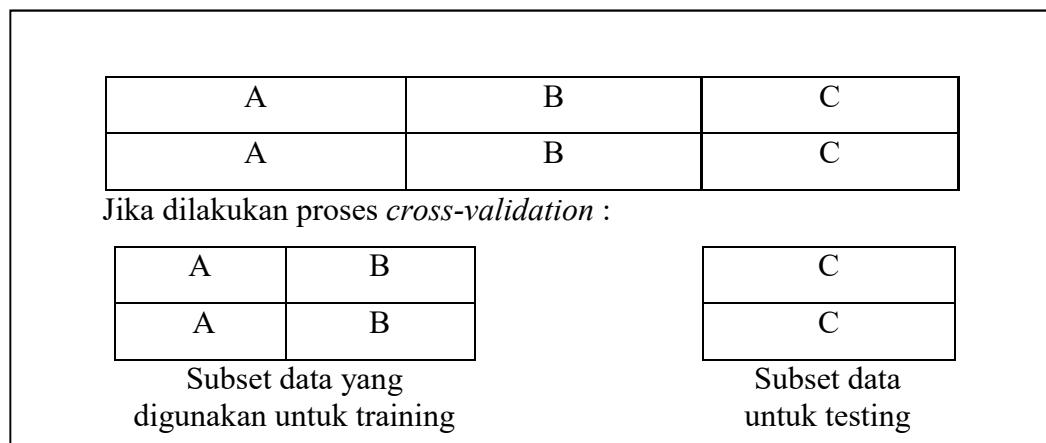
nilai minimumnya adalah $\min(x_i)$ maka untuk mengubah data ke skala yang baru menggunakan persamaan 2.1.

$$\hat{x} = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (2.1)$$

2.1.2 Cross-Validation

Cross-Validation sangat penting dalam rangka menemukan parameter model yang terbaik. [14]. Dalam cross-validation, data dibagi ke dalam k sampel dengan ukuran yang sama. Jika menggunakan $k-1$ sample untuk training, maka sisanya untuk data testing. Pembagian seperti ini disebut juga dengan k -fold. Dengan k adalah konstanta yang membagi jumlah data sebanyak k subsample dan dilakukan pengujian sebanyak k -kali. Misalkan $k = 10$, maka ada 10 subset data. Jika menggunakan 9 subset untuk data training maka 1 subset sisanya untuk data testing.

Contohnya pada Gambar 2.1, terdapat data dengan subset A, B dan C dan $\text{cross validation} = 3$, jika C dijadikan subset data testing maka subset A dan B menjadi training data. Dilakukan untuk semua kemungkinan kombinasi subset dan dilakukan pengujian besarnya error model tersebut. Model yang memberikan error terkecil adalah model dengan kombinasi yang terbaik.



Gambar 2.1 Cross-validation dengan $k=3$

2.1.3 Metode Evaluasi

Metode evaluasi dari model regresi yang biasa digunakan adalah

2.1.3.1 Explained variance score.

Jika \hat{y} adalah keluaran model estimasi dan y adalah target aktual maka Var adalah variance dari kuadrat standard deviasi yang menggunakan persamaan 2.2. Standard deviasi adalah nilai statistik yang digunakan untuk menentukan bagaimana sebaran data dalam sampel.

$$\text{explained_variance}(y, \hat{y}) = 1 - \frac{Var(y - \hat{y})}{Var(y)} \quad (2.2)$$

Di dalam model regresi, nilai $\text{explained_variance}$ yang terbaik adalah sama dengan satu jika kurang dari satu maka model tersebut belum optimal.

2.1.3.2 Mean absolute error

Mean absolute error (MAE) jika \hat{y} adalah keluaran model estimasi, y adalah target aktual dari $n_{samples}$, persamaan MAE ditunjukkan didalam persamaan 2.3.

$$MAE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} |y_i - \hat{y}_i| \quad (2.3)$$

2.1.3.3 Median Absolute Error

Median Absolute Error(MedAE) merupakan model evaluasi yang cukup menarik karena dapat digunakan untuk data outlier. Nilai loss dihitung dari nilai median semua nilai absolute selisih data target dan data hasil prediksi. jika \hat{y} adalah keluaran model estimasi dari i sample, y adalah target aktual dari $n_{samples}$, persamaan 2.4 menunjukkan MedAE.

$$MedAE(y, \hat{y}) = \text{median}(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|) \quad (2.4)$$

2.1.3.4 Mean Square Error atau MSE

Metrik risk yang berkaitan dengan nilai yang diharapkan dari kuadrat error loss. Jika \hat{y} adalah keluaran model estimasi dari I_{sample} dan y adalah target

aktual maka nilai MSE dari $n_{samples}$, maka model evaluasi tersebut ditunjukkan oleh persamaan 2.5

$$MSE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2 \quad (2.5)$$

2.1.3.5 Koefisien Determinasi R^2

Menghitung koefisien determinasi dari model regresi. Nilai ini akan mengukur seberapa baik performansi model dalam memprediksi data test. Nilai terbaik dari R^2 adalah 1.0, nilai R^2 bisa bernilai negative yang berarti model yang dihasilkan belum optimal. Model konstan yang selalu memprediksi nilai y , mengabaikan input, maka nilai R^2 bernilai 0. Jika \hat{y} adalah keluaran model estimasi dari i sample, y adalah target aktual dari $n_{samples}$ maka nilai R^2 ditunjukkan di dalam persamaan 2.6 dan 2.7.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{samples}-1} (y_i - \bar{y})^2} \quad (2.6)$$

dimana :

$$\bar{y} = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} y_i \quad (2.7)$$

2.2 Indeks Harga Konsumen

Inflasi merupakan suatu kecenderungan meningkatnya harga-harga barang dan jasa secara umum dan terus menerus. Inflasi yang rendah dan stabil merupakan prasyarat untuk mewujudkan kesejahteraan masyarakat. Menyadari perlunya pengendalian inflasi dalam suatu negara, pemerintah memerlukan kerjasama dan koordinasi lintas instansi, yakni antara Bank Indonesia dengan Pemerintah baik di tingkat pusat maupun daerah. Pemantauan inflasi ini tiap waktu harus dilakukan secara cermat dengan melihat indikator-indikator perubahan harga pada komoditas tertentu. Perkembangan kenaikan harga

sejumlah barang dan jasa secara umum dalam suatu periode waktu ke waktu biasanya dinyatakan dalam bentuk persen dan dikenal dengan laju inflasi.

Perhitungan laju inflasi yang umum digunakan menggunakan dua cara yaitu Indeks Harga Konsumen (IHK) dan GDP Deflator (Gross Domestic Product). Perhitungan laju inflasi tahun ke n menggunakan persamaan 2.8.

$$\text{inflasi}_n = \frac{(IHK_n - IHK_{(n-1)})}{IHK_{n-1}} \times 100\% \quad (2.8)$$

Dimana :

IHK_n : IHK pada tahun dasar

IHK_{n-1} : IHK pada tahun sebelumnya

Indikator yang paling sering digunakan untuk menganalisa dan mengukur laju inflasi adalah IHK (Indeks Harga Konsumen) [11]. Indeks Harga Konsumen (IHK) merupakan salah satu indikator ekonomi yang memberikan informasi mengenai harga barang dan jasa yang dibayar oleh konsumen. Perhitungan IHK dilakukan untuk mengetahui perubahan harga beli di tingkat konsumen yang pada umumnya dikonsumsi oleh masyarakat. IHK menghitung harga rata-rata dari barang dan jasa yang dikonsumsi oleh rumah tangga (household). Jenis barang dan jasa tersebut dikelompokkan menjadi 7 kelompok, yaitu Bahan makanan; Makanan jadi, minuman, rokok, dan tembakau; perumahan, Air, Listrik, Gas dan Bahan Bakar; Sandang; Kesehatan; Pendidikan, rekreasi dan olahraga; Transportasi, komunikasi dan jasa keuangan. Kelompok tersebut sudah sangat cukup mewakili “harga-harga secara umum”, oleh karena itu IHK paling sering digunakan untuk menghitung inflasi.

Perhitungan IHK di Indonesia dilakukan oleh Badan Pusat Statistik (BPS). Sejak Januari 2014 sampai dengan saat ini, data IHK diperoleh dari survey yang dilakukan di 82 kota di seluruh Indonesia. Sebelumnya, data IHK hanya diperoleh dari 66 kota di Indonesia. Di Jawa timur, IHK dihitung di 8 kota/kabupaten perwakilan yaitu Surabaya, Kota Malang, Kota Kediri, Kota Madiun, Kota Probolinggo, Kab Jember, Kabupaten Sumenep dan Kab

Banyuwangi. Sebelum tahun 2014, IHK di Jawa Timur hanya dihitung dari tujuh kota, Kabupaten Banyuwangi belum dijadikan sampling perhitungan BPS.

Perhitungan IHK didasarkan pada survey Biaya Hidup (SBH) 2012. Data tersebut digunakan sebagai dasar untuk menentukan paket komoditas, penimbang, cakupan kota dan tahun dasar dalam pengolahan IHK. Oleh karena itu, setiap lima tahun, diadakan perbaikan daftar timbang sehingga nilai Indeks Harga Konsumen berada pada skala yang sama yaitu 100. Perhitungan IHK ini dilakukan oleh BPS setiap bulan dengan menggunakan metode Modified Laspeyres seperti persamaan 2.9.

$$I_n = \frac{\sum_{i=1}^k \frac{P_{ni}}{P_{(n-1)i}} (P_{(n-1)i} \cdot Q_{0i})}{\sum_{i=1}^k P_{0i} \cdot Q_{0i}} \times 100 \quad (2.9)$$

Dimana :

- I_n : Indeks bulan ke-n
- P_{ni} : Harga jenis komoditi i pada bulan ke-n
- $P_{(n-1)i}$: Harga jenis komoditi i pada bulan ke n-1
- $P_{(n-1)i} \cdot Q_{0i}$: Nilai konsumsi jenis komoditi i bulan ke n-1
- $P_{0i}Q_{0i}$: Nilai konsumsi jenis komoditi i pada tahun dasar

Pengumpulan data untuk perhitungan IHK ini dilakukan dengan frekuensi yang berbeda. Misalkan pengumpulan data harga beras dilakukan secara harian di Jakarta dan dilakukan mingguan di kota-kota lainnya. Beberapa komoditas yang termasuk ke dalam kebutuhan pokok, data harganya dikumpulkan setiap minggu yaitu pada hari Senin dan Selasa. IHK harian merupakan perhitungan rata-rata dari harga harian hasil survey. Prediksi yang dilakukan oleh Bank Indonesia menetapkan standard deviasi sekitar 1% untuk laju inflasi.

2.3 Support Vector Regression

Support Vector Regression (SVR) merupakan bagian dari metode Support Vector Machine (SVM). Dalam kasus klasifikasi, output dari model SVM

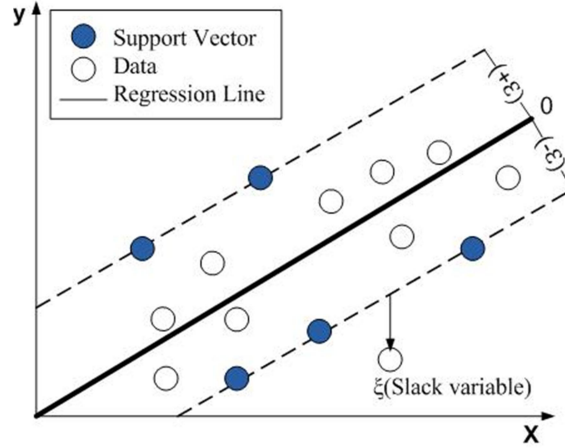
berupa bilangan bulat atau diskrit. Sedangkan di dalam kasus regresi, output datanya berupa bilangan riil atau kontinu. Dengan menggunakan konsep *e-insensitive loss function*, yang diperkenalkan oleh Vapnik, SVM bisa digeneralisasi untuk melakukan pendekatan regresi. SVM merupakan metode yang dapat mengatasi masalah *overfitting* dibandingkan dengan *neural network*, sehingga akan menghasilkan model prediksi yang lebih bagus [14]. SVR merupakan bagian dari SVM. Support Vector Regression berdasarkan pada teori Structural Risk Minimization untuk mengestimasi suatu fungsi dengan cara meminimalkan batas atas dari generalization error sehingga SVR bisa mengatasi *overfitting*.

Tujuan dasar Support Vector Machine untuk klasifikasi adalah mencari sebuah hyperplane pemisah yang memaksimalkan margin antara dua kelas. Dipastikan juga data baru yang akan dimasukkan dalam prediksi mempunyai klasifikasi yang lebih baik atau kondisi ini bisa disebut generalisasi yang lebih baik. Model yang dihasilkan oleh klasifikasi Support Vector yaitu dari SVM hanya tergantung pada subset dari data pelatihan, karena fungsi cost untuk membuat model tidak memperdulikan data pelatihan yang terletak melampaui margin.

Dalam kasus regresi, tujuan dasar SVR adalah untuk menemukan fungsi $f(x)$ yang memiliki paling banyak ϵ deviasi dari target sebenarnya yang diperoleh dari semua data training, dan pada saat yang sama fungsi tersebut harus sedatar mungkin. Dengan kata lain, error tidak dipedulikan, selama error kurang dari epsilon ϵ . Di dalam SVM ataupun SVR dikenal support vector, support vector adalah data latih yang digunakan dalam pengujian.

Misalkan ada n -data training (x_i, y_i) dengan $i = 1, \dots, n$, dengan menggunakan metode SVR akan ditemukan suatu fungsi $f(x)$ yang mempunyai deviasi atau epsilon, ϵ , dari target aktual y untuk semua data training. Di dalam SVR, akan diusahakan semua data masuk di dalam tube. Ditunjukkan di dalam Gambar 2.1, gambar ini menunjukkan *e-insensitive loss function* dari Support Vector Regression. Dimana ϵ merupakan batas dari daerah yang dikenal dengan

tube. Jika ε bernilai 0 maka regresi bernilai sempurna atau digambarkan garis yang berada di tengah.



Gambar 2.2 Arsitektur Support Vector Machine Regression

Semua titik atau data yang berada di luar area tube akan dikenai nilai penalti atau dikenal dengan parameter C pada SVR. Titik yang keluar dari rentang epsilon ε , bisa ditambahkan variabel slack atau ξ , hal ini juga untuk mengatasi masalah pembatas yang tidak layak dalam problem optimasi.

Secara matematika dapat dijabarkan, misalnya mempunyai persamaan 2.10 sebagai garis regresi.

$$f(x) = w^T \varphi(x) + b \quad (2.10)$$

Dimana $\varphi(x)$ menunjukkan suatu titik di dalam feature space F hasil pemetaan x di dalam input space.

Koefisien w dan b diestimasi dengan cara meminimalkan fungsi resiko (risk function) yang didefinisikan dalam persamaan 2.11

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\lambda} \sum_{i=1}^{\lambda} L_e(y_i, f(x_i)) \quad (2.11)$$

Subject to

$$y_i - w\varphi(x_i) - b \leq \varepsilon$$

$$w\varphi(x_i) - y_i + b \leq \varepsilon, i = 1, 2, \dots, \lambda$$

Dimana

Faktor $\|w\|$ dinamakan reguralisasi. Meminimalkan nilai $\|w\|^2$ akan membuat fungsi setipis mungkin sehingga bisa mengontrol kapasitas fungsi.

Faktor kedua dalam fungsi tujuan adalah kesalahan empirik (empirical error) yang diukur dengan ϵ -insensitive loss function harus meminimalkan norm dari w agar mendapatkan generalisasi yang baik untuk fungsi regresi f .

Karena itu perlu menyelesaikan problem optimasi didalam persamaan 2.12

$$\min \frac{1}{2} \|w\|^2 \quad (2.12)$$

Subject to

$$y_i - w\varphi(x_i) - b \leq \epsilon$$

$$w\varphi(x_i) - y_i + b \leq \epsilon, i = 1, 2, \dots, \lambda$$

$$\begin{aligned} \sum e^2 &= \sum (y - \hat{y})^2 \\ &= \sum (y - (b_0 + b_1 x))^2 \end{aligned} \quad (2.13)$$

Asumsikan bahwa ada suatu fungsi f yang dapat mengaproksimasi semua titik (x_i, y_i) dengan presisi ϵ . Dalam kasus ini diasumsikan bahwa semua titik ada dalam rentang $f \pm \epsilon$ (feasible). Dalam hal constraint yang tidak layak (infeasible), dimana mungkin ada beberapa titik yang mungkin keluar dari rentang $f \pm \epsilon$ (feasible), bisa ditambahkan variabel slack ξ, ξ^* untuk mengatasi masalah pembatas yang tidak layak (infeasible constraint) dalam problem optimasi.

Selanjutnya problem optimasi di atas bisa diformulasikan seperti persamaan 2.14

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{\lambda} \sum_{i=1}^{\lambda} (\xi_i, \xi_i^*) \quad (2.14)$$

Subject to

$$y_i - w^T \varphi(x_i) - b - \xi_i \leq \epsilon, i=1, 2, \dots, \lambda$$

$$w\varphi(x_i) - y_i + b - \xi_i^* \leq \epsilon, i = 1, 2, \dots, \lambda$$

$$\xi_i, \xi_i^* \geq 0$$

Konstanta $C > 0$ menentukan tawar menawar (trade off) antara ketipisan fungsi f dan batas atas deviasi lebih dari ϵ masih ditoleransi. Semua deviasi lebih besar daripada ϵ akan dikenakan penalty sebesar C .

Dalam SVR, ϵ ekuivalen dengan akurasi dari aproksimasi terhadap data training,

- Nilai ϵ yang kecil terkait dengan nilai yang tinggi pada variable slack ξ_i^* dan akurasi aproksimasi yang tinggi.

- Nilai ε yang tinggi berkaitan dengan nilai ξ_i^* yang kecil dan aproksimasi yang rendah.
- Menurut persamaan (2.14) nilai yang tinggi untuk variabel slack akan membuat kesalahan empirik mempunyai pengaruh yang besar terhadap faktor regulasi.

Dalam SVR, *Support Vector* adalah data training yang terletak pada dan luar batas f dari fungsi keputusan, karena itu jumlah *support vector* menurun dengan naiknya ε

Dalam formulasi dual, problem optimasi dari SVR adalah sebagai berikut :

$$\max -\frac{1}{2} \sum_{i=1}^{\lambda} \sum_{j=1}^{\lambda} (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K \langle x_i, x_j \rangle + \sum_{i=1}^{\lambda} (\alpha_j - \alpha_j^*) y_i - \varepsilon \sum_{i=1}^{\lambda} (\alpha_j + \alpha_j^*) \quad (2.15)$$

Subject to

$$\sum_{i=1}^{\lambda} (\alpha_j - \alpha_j^*) = 0$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \dots, \lambda$$

$$0 \leq \alpha_i^* \leq C, i = 1, 2, \dots, \lambda$$

Dimana C didefinisikan oleh user, $K \langle x_i, x_j \rangle$ adalah dot- product kernel yang didefinisikan sebagai $K \langle x_i, x_j \rangle = \varphi^T(x_i) \varphi(x_j)$

Dengan menggunakan *langrange multiplier* dan kondisi optimalitas, fungsi regresi secara eksplisit ditunjukkan didalam persamaan 2.16

$$f(x) = \sum_{i=1}^{\lambda} (\alpha_j - \alpha_j^*) K \langle x_i, x_j \rangle + b \quad (2.16)$$

Performansi model prediksi SVR tergantung pada penggunaan hyperparameters C , epsilon (ε), degree, parameter kernel (gamma) dan fungsi kernel yang digunakan.

C : parameter C untuk memberikan nilai penalty pada data jika terjadi error pada prediksi

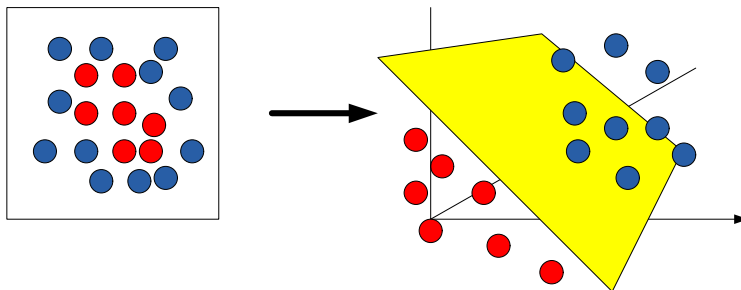
epsilon : parameter epsilon menandakan epsilon tube yang berarti tidak ada penalti di dalam loss function terhadap point prediksi selama masih berada di dalam jarak epsilon dari nilai aktual

degree : degree hanya digunakan oleh fungsi kernel polynomial (poly). Nilai parameter degree tidak digunakan pada kernel lainnya

gamma : koefisien dari kernel RBF, Poly

Fungsi Kernel

Pada umumnya, data yang kompleks membutuhkan fungsi yang lebih ekspresif daripada hanya linear. Di dalam kasus non linear, SVR akan menstranformasikan original *input space* ke dalam dimensi yang lebih tinggi, *feature space*. Setelah itu, SVR akan melakukan perhitungan sama seperti perhitungan linear, untuk mencari hyperplane regresi yang optimal.



Gambar 2.3 Kernel mapping dari input space ke feature space

Gambar 2.3 yaitu gambar *mapping kernel*, data warna biru dengan kelas 1 dan data warna merah dengan kelas 0. Artinya, menunjukkan bagaimana data warna biru dan merah dari input space tidak bisa dipisahkan, sehingga perlu di transformasikan ke dimensi yang lebih tinggi, *feature space* menggunakan fungsi kernel.

Tabel 2.1 Fungsi Kernel

Kernel	Formula
Linear	(x, x')
RBF / Gaussian	$\exp(-\gamma x - x' ^2)$ γ adalah gamma yang nilainya harus lebih besar dari nol
Polynomial	$(\gamma(x, x') + r)^d$ d adalah degree, sedangkan r adalah coef0

Dalam penelitian ini fokus pada penggunaan tiga kernel yaitu Linear, RBF dan Poly. Persamaan matematika dari fungsi kernel ditunjukkan di dalam Tabel 2.1, fungsi kernel.

2.4 Cloud Computing

Salah satu definisi dan standarisasi NIST (National Institute of Standard and Technology), cloud computing merupakan model yang memungkinkan penggunaan sumber daya (resource) secara bersama sama dan mudah, menyediakan jaringan akses dimana mana, dapat dikonfigurasi, dan layanan yang digunakan sesuai keperluan. Di dalam penelitian yang ditulis (Prasad & Atukuri, 2012) disebutkan lima fitur utama dari cloud computing :

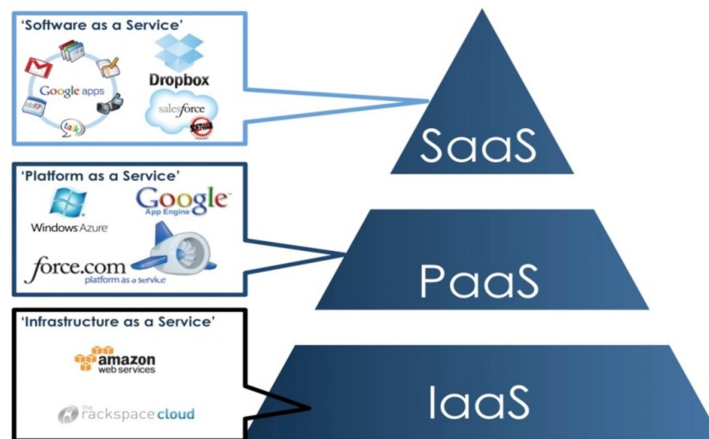
1. Service demand on self : pelanggan dapat dengan mudah mengakses fasilitas komputasinya seperti server , database jika diperlukan
2. Akses Ubiquitous. Fasilitas dapat ini memudahkan akses pengguna dari beberapa platform dan media akses misalkan mediakomputer desktop, laptop ataupun mobile phone
3. Location-independent Resource pooling adalah sumber daya cloud computing ini dapat disediakan dan digunakan secara bersama sama meskipun pengguna dan lokasi layanan berada pada lokasi yang berbeda. Pengguna tidak perlu tahu dimana lokasi fisik dari server cloud yang diaksesnya tetapi cukup mengetahui bagaimana kebutuhan mereka dapat dipenuhi
4. Rapid Elasticity. Sumber daya sistem dapat tersedia dengan cepat sesuai dengan kebutuhan pelanggan (elastis). Dengan kata lain layanan dapat selalu diperbarui dan diperbaiki dan dapat diakses oleh pengguna.
5. Measured Service Penggunaan sumber daya dapat dimonitor , control dan dilaporkan. Pengguna dapat membayar sesuai dengan layanan yang disewa dan digunakan.

Pada teknologi Cloud Computing, terdapat tiga model layanan dasar yang dapat dipilih sesuai dengan kebutuhan meliputi IAAS, PAAS dan SAAS,

pembagian jenis layanan ini dilakukan dengan tujuan untuk membantu menyesuaikan keperluan dari pengguna.

IAAS (Infrastructure as Service)

IaaS merupakan jenis layanan pada cloud computing yang fokus menyediakan sarana jaringan komputer, perangkat server, media penyimpanan beserta dengan proses virtualisasinya. Pada IAAS pengguna seperti memiliki server sendiri, walaupun secara fisik tidak ada atau hanya Virtual Machine (VM). Contoh Penyedia layanan ini adalah Amazon Elastic Compute Cloud (EC2), RackSpace Cloud. Di Indonesia ada Telkomcloud dan Cloud Server Indonesia



Gambar 2.4 Layanan Cloud beserta Penyediaanya

PAAS (Platform As A service)

merupakan jenis layanan pada Cloud Computing yang menekankan kepada penyediaan platform untuk membantu pengembangan perangkat lunak secara cepat dan mudah. Penyedia PaaS bertanggungjawab dalam infrastruktur cloud, sistem operasi dan lisensi platform. Layanan platform umumnya berbasis web dimana banyak tersedia fitur yang memudahkan programmer mengembangkan perangkat lunaknya. Contoh penyedia layanan PaaS : Windows Azure, force.com, Google App Engine,

SAAS (Software As A Service)

Jenis layanan yang diberikan cloud computing model ini adalah bentuk pemakaian perangkat lunak secara bersama sama. Umumnya perangkat lunak

berbasis web. SaaS merupakan jenis layanan cloud Computing yang paling banyak digunakan. Pengguna tidak perlu tahu sulitnya instalasi dan konfigurasi aplikasi. Cukup dengan mengakses aplikasi melalui browser dan koneksi jaringan komunikasi atau internet. Contoh penyedia layanan SaaS ini adalah Google Apps, Salesforce.com, DropBox, Amazon Web Service dan Azure.

Model layanan cloud digambarkan seperti berikut

Cloud Service Models

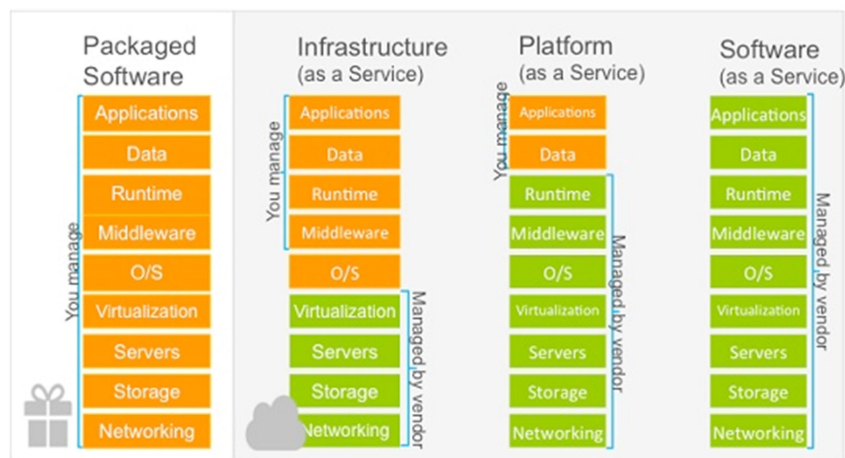


Figure 1.

Source: Microsoft Azure

Gambar 2.5 Model Layanan Cloud

Ditunjukkan perbedaan antara beberapa jenis layanan. Pada model konvensional, misalkan di dalam instansi Pemerintahan, yang mempunyai infrastruktur IT bertanggung jawab terhadap Jaringan, Penyimpanan, OS, Runtime, Data, Application. Jika di dalam instansi tersebut tidak ada administrator sistem, maka jika terjadi gangguan terhadap sistem tersebut sangat mempengaruhi jalannya aplikasi. Berbeda dengan pelayanan di cloud. Model layanan Infrastructure, *IaaS* jaringan, virtualisasi, server, penyimpanan merupakan tanggung jawab penyedia cloud. Pengguna seperti mempunyai server sendiri. Pengguna bisa menginstall OS sesuai dengan kebutuhan mereka, membangun aplikasinya sendiri dan manage data sendiri. Model Layanan Platform, *PaaS* pada layanan ini pengguna dapat mengembangkan perangkat lunak sendiri dan mengelola data sendiri. Tidak perlu memikirkan instalasi Sistem operasi yang akan digunakan apalagi pemeliharaan server. Model layanan *PAAS* ini menjadi

tren untuk pengembangan perangkat lunak. Di model layanan *SaaS* pengguna tidak perlu memikirkan pengembangan perangkat lunaknya atau cara mengelola datanya. Semua disediakan dan dikelola oleh penyedia cloud. Pengguna hanya menikmati layanan yang diberikan. Aplikasi office online dari microsoft merupakan bentuk layanan ini, layanan ERP, CRM (Customer Relation Manager), email merupakan layanan yang biasanya diberikan di model SaaS

2.4.1 Penyedia layanan cloud computing

Amazon Web Service

Amazon Web Service adalah sekumpulan layanan berbasis cloud computing yang dirilis perdana sejak tahun 2006. Pada tahun 2006 Amazon merilis Amazon Elastic Compute Cloud atau yang dikenal dengan Amazon EC2. Amazon EC2 atau Amazon Elastic Compute Cloud merupakan salah satu penyedia layanan infrastruktur *cloud computing* komersial terintegrasi yang diperuntukkan untuk individu atau perusahaan. Amazon EC2 berbentuk platform komputasi berupa virtual komputer yang dapat dikustomisasi maupun dikembangkan dengan menggunakan prinsip cluster dan load balance [13]. Inilah awal dari layanan IaaS. Selain menyediakan layanan IaaS, Amazon menyediakan service lainnya, aplikasi berbayar. Aplikasi berbayar tersebut tersedia di AWS Marketplace. Pengguna dapat memilih aplikasi tersebut dan membayarnya berdasarkan kebutuhan. Amazon menyediakan beberapa tipe layanan berdasarkan hardware yang terinstall sebagai berikut :

Tabel 2.2 Beberapa contoh Tipe Layanan Amazon

Type	CPU	Memory
T2.nano	1	0.5
T2.micro	1	1
T2.small	1	2
T2.medium	2	4
T2.large	2	8
T2.xlarge	4	16

Pengguna dapat memilih layanan yang tersedia, membayar layanan tersebut sesuai dengan kebutuhan pengguna.

Microsoft Azure

Microsoft Azure dikenalkan pada tahun 2010. Sama halnya dengan Amazon, Microsoft Azure menjadi perusahaan besar untuk penyedia Cloud Computing. Layanan yang ditawarkan keduanya juga hampir sama.

Menurut penelitian yang dilakukan oleh T. Madhuri dan P. Sowjanya, [16] dalam studinya dibandingkan layanan dari Microsoft Azure dan Amazon AWS Cloud. Di dalam Tabel 2.3 menjelaskan perbandingan antara Cloud penyedia amazon dengan Azure.

2.4.2 Cara mengakses Cloud service

Layanan cloud dapat diakses dengan cara yang berbeda beda. Untuk layanan cloud IaaS atau Virtual Machine, biasanya diakses dengan menggunakan Secure Shell atau SSH. SSH adalah protokol jaringan yang memungkinkan pertukaran data melalui saluran yang aman antara dua perangkat jaringan. Terutama banyak digunakan pada sistem berbasis Linux dan Unix untuk mengakses akun shell, SSH dirancang sebagai pengganti Telnet dan shell remote tak aman lainnya, yang mengirim informasi, terutama kata sandi, dalam bentuk teks sederhana yang membuatnya mudah untuk dihalangi. Enkripsi yang digunakan oleh SSH menyediakan kerahasiaan dan integritas data melalui jaringan yang tidak aman seperti Internet.

SSH menggunakan kriptografi kunci publik untuk mengotentikasi komputer remote dan biarkan komputer remote untuk mengotentikasi pengguna, jika perlu. SSH biasanya digunakan untuk login ke mesin remote dan mengeksekusi berbagai perintah, tetapi juga mendukung tunneling, forwarding TCP port dan X11 connections; itu dapat mentransfer file menggunakan terkait SFTP atau SCP protocols. SSH menggunakan client-server model. Port 22 merupakan port standard yang ditetapkan untuk menghubungi server SSH. Sebuah klien program SSH ini biasanya digunakan untuk membangun koneksi ke SSH daemon untuk dapat diremote. Keduanya biasanya terdapat pada sistem operasi modern, termasuk Mac OS X, Linux, FreeBSD, Solaris dan OpenVMS.

Tabel 2.3 Perbandingan Amazon Web Service dan Microsoft Azure

No	Amazon Web Service	Microsoft Azure
1	Biaya awal EC2 sangat murah sekitar \$40 tiap bulan. Dengan ini pengguna bisa mempunyai instance (virtual machine) yang berjalan dengan sistem operasi windows 2008 R2 Server	EC2 lebih murah daripada Azure, tetapi Azure menawarkan zero maintenance. Pengguna hanya perlu implementasi aplikasinya dan Azure yang akan melakukan maintenace terhadap aplikasi ini.
2	EC2 mudah dipahami. Konsep dari EC2 sangat sederhana bagi pengguna yang tidak terbiasa menjalan software secara virtual.	Scalability Azure lebih baik daripada amazon. Perubahan bisa sangat mudah dilakukan jika terjadi penambahan user dan konsumsi data yang besar
3	Mesin bisa diakses secara individu	Mesin tergabung dalam group cloud service dan terkait dengan domain yang sama tetapi berbeda port
4	Media penyimpanan EBS cukup cepat untuk kebutuhan Big data	Media penyimpanan standard cukup sulit untuk Big Data. Sehingga diperlukan media penyimpanan premium
5	Kematangan untuk pemrosesan Big Data lebih baik	Kematangan untuk pemrosesan Big Data masih kurang
6	Mendukung machine learning tool	Mendukung machine learning tool di azure
7	Dari sisi kecepatan, Azure mempunyai kecepatan 56% persen lebih cepat dibandingkan Amazon S3	
8	Availability : Azure mempunyai response time lebih cepat 25% daripada Amazon S3	

Untuk transfer data, Cloud memperbolehkan penggunaanya mengakses service FTP di port 21. Penyedia cloud juga menyediakan SaaS, sehingga tidak perlu menggunakan SSH dan cara akses SaaS biasanya lebih mudah dipahami.

2.4.3 Migrasi aplikasi ke dalam Cloud Computing

Organisasi perlu menyusun strategi dalam melaksanakan migrasi aplikasinya ke dalam Cloud Computing. Strategi ini diperlukan karena aplikasi akan dipindahkan ke dalam lingkungan komputasi yang berbeda, jadi harus ada panduan atau kerangka kerja yang dilakukan untuk mempertimbangkan segala kemungkinan yang terjadi. Perlu dipertimbangkan masalah administrator, biaya, pemilihan service cloud, pemilihan penyedia cloud, resiko yang dihadapi dan beberapa kemungkinan lainnya. Beberapa strategi dan metodologi berkaitan dengan migrasi ke cloud dibahas oleh Pappas. Pappas melakukan beberapa review metodologi dan tool untuk melakukan migrasi ke Cloud [17]. Salah satu metodologi yang dibahas adalah :

Migrasi Cloud-RMM. Proses yang dilakukan di dalam migrasi ke dalam Cloud adalah :

Tabel 2.4 Framework migrasi Cloud-RMM

No	Proses	Tugas
1	Perencanaan	Studi kelayakan, analisa kebutuhan migrasi, pemilihan penyedia cloud dan layanan cloud, strategi migrasi
2	Pelaksanaan	Modifikasi kode pemrograman aplikasi, Ekstrasi arsitektur, Ekstrasi data, Trasformasi
3	Evaluasi	Pengembangan, Pengujian, Validasi
4	Crosscutting concerns	Tata kelola, Keamanan, Pelatihan, Estimasi biaya yang dikeluarkan, Manajemen perubahan, Multitenancy

2.5 Prediksi Finansial dengan menggunakan high frequency data

Kondisi perekonomian, pasar keuangan sering merupakan permasalahan yang sangat kompleks. Beberapa prediksi atau regresi berkaitan dengan masalah keuangan banyak dikembangkan untuk memperkirakan kondisi perekonomian ini. Hal ini dilakukan untuk perencanaan kebijakan moneter yang lebih baik. Saat ini perkembangan informasi yang berkaitan dengan keuangan sangat banyak, oleh karena itu pelaku pasar harus menyadari semua kemungkinan yang mungkin bisa terjadi berkaitan dengan keuangan dengan menggunakan informasi ini. Potensi pola yang tersembunyi dalam data mungkin bisa memberikan peringatan dini untuk kejadian, misalkan inflasi yang tinggi.

Pemanfaatan prediksi IHK menggunakan data harian merupakan prediksi dengan menggunakan *high frequency* data atau data yang keberadaannya lebih sering. Beberapa penelitian telah membahas masalah prediksi ini diantaranya : Di dalam penelitian yang dilakukan Ekaterina Zankova, dibahas prediksi harga secara real time [12]. Prediksi dilakukan dengan menggunakan high frequency data dan 4 metode regresi, yaitu Decision Tree Regressor, Multilayer Perceptron Regressor , K-Nearest Neighbour Regressor dan Support Vector Regression menggunakan kernel linear. Karena penelitian ini menfokuskan prediksi secara real time, pengalokasian data training dan testing mendapatkan perhatian khusus. Di dalam penelitian ini pengalokasi data menggunakan metode sliding window. Hasil dari penelitian ini adalah metode SVR mempunyai akurasi yang bagus tetapi jika melihat juga dari segi waktu maka Multilayer perceptron mempunyai performansi yang lebih baik.

Di dalam penelitian mroeconomy Forecasting with mixed frequency data [13] dibahas banyak seri makroekonomi seperti pertumbuhan output riil AS mengambil sampel kuartalan, meskipun prediksi berpotensi diamati menggunakan frekuensi yang lebih tinggi. Pendekatan mixed data-frequency Sampling (MIDAS) memanfaatkan data bulanan untuk memprediksi kuartalan. Dan menghasilkan bahwa penggunaan data bulanan pada kuartal mengarah ke peningkatan dalam peramalan pertumbuhan output kuartal. Jika semua variabel dalam model harus dijadikan sampel pada frekuensi yang sama, misalnya, data yang tersedia pada frekuensi bulanan harus dikonversi ke frekuensi kuartalan,

dengan menggunakan rata-rata data bulan (atau mengambil nilai bulan terakhir di kuartalan). Jadi sangat dimungkinkan untuk membangun prediksi dengan menggunakan data harian untuk memperbaiki perhitungan bulanan

2.6 Linear Regression

Regresi linier adalah algoritma pembelajaran supervised yang menghasilkan prediksi output dengan memetakan hubungan linier yang optimal antara data output dan input. Data output merupakan variabel yang akan diprediksi disebut variable dependen $\rightarrow y$ sedangkan data input yang mendasari prediksi atau disebut prediktor atau variable independen $\rightarrow x$

Model Regresi Linear sederhana adalah ditunjukkan oleh persamaan 2.17

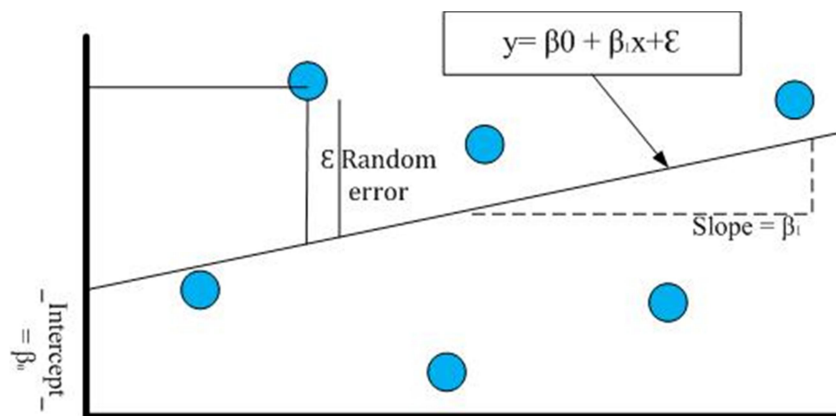
$$\hat{y} = b_0 + b_1 x \quad (2.17)$$

dimana :

\hat{y} = Variabel Response atau Variabel Akibat (Dependent)

x = Variabel Predictor atau Variabel Faktor Penyebab (Independent)

b_0 b_1 = koefisien



Gambar 2.6 Arsitektur Linear Regression

Gambar 2.6 menunjukkan arsitektur Linear Regression, data digambarkan oleh titik berwarna biru. Dan garis yang melintasinya adalah garis regresi yang mewakili persamaan 2.17. Tujuan dari analisa regresi linear sederhana ini untuk mencari nilai yang optimal dari b_0 dan b_1 sehingga digunakanlah pendekatan least square.

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.18)$$

Tujuan dari pendekatan ini yaitu berusaha meminimumkan jumlah error kuadrat (SSE) dari semua data yang digunakan untuk menemukan estimasi garis regresi.

Maka koefisien b_1 dan b_0 :

$$b_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (2.19)$$

$$b_0 = \bar{y} - b_1 \bar{x} \quad (2.20)$$

2.7 Kernel Ridge Regression (KRR)

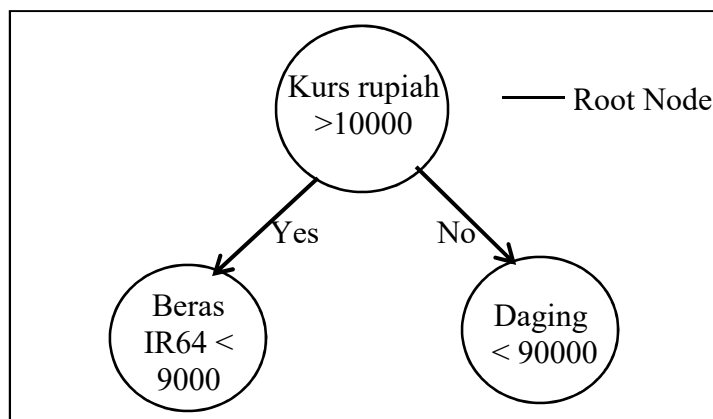
Metode ini digunakan dalam penelitian ini sebagai pembanding metode Support Vector Regression (SVR) untuk mencari nilai estimasi IHK harian. Di dalam model regresi linear tujuannya meminimalkan nilai SSE. Jika hanya ada satu nilai x , umumnya model bekerja dengan baik, namun bila ada banyak nilai x , hal itu dapat menyebabkan *over-fitting*. Untuk menghindari hal ini, bisa menggunakan Ridge Regression, di mana algoritma regresi menambahkan istilah regularisasi ke SSE biasa disebut parameter α . Regularisasi membantu mencapai keseimbangan akurasi dan kesederhanaan dalam fungsinya.

Ridge Regression merupakan salah satu model linear yang mempunyai regularisasi l_2 -norm. Kernel Ridge Regression merupakan pengembangan dari metode Ridge Regression. Di dalam metode Kernel Ridge Regression digabungkan metode Ridge Regression dengan trik kernel. Penggabungan ini bertujuan agar Ridge Regression dapat mengatasi permasalahan non linier. Karena hampir semua masalah berupa non linear. Kernel Ridge ini hampir mirip dengan SVR linear, perbedaannya, KRR menggunakan *Squared Error Loss* sedangkan SVR menggunakan *ϵ -insensitive loss*, keduanya dikombinasikan dengan regularisasi l_2 . Kelebihan metode Kernel Ridge, biasanya proses fitting metode ini mempunyai waktu yang lebih cepat dibandingkan SVR. Di dalam metode Kernel Ridge Regression mempunyai parameter α dan γ , α memberikan penalty kepada data jika data keluar dari batas yang ditentukan.

2.8 Random Forest Regressor

Disebut dengan nama *Forest* karena dalam metode ini merupakan kumpulan tree (decision tree). Variable data di dalam Random forest di acak kemudian dibentuklah tree – tree yang mempunyai rule sendiri.. *Decision Tree* adalah sebuah diagram alir yang berbentuk struktur pohon yang mana setiap internal node menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pengujian tersebut dan leaf node menyatakan nilai kontinu atau distribusi kelas. Node yang paling atas disebut sebagai root node atau node akar. Sebuah root node akan memiliki beberapa edge keluar tetapi tidak memiliki edge masuk, internal node akan memiliki satu edge masuk dan beberapa edge keluar, sedangkan leaf node hanya akan memiliki satu edge masuk tanpa memiliki edge keluar.

Pada continuous attribute, penyelesaian *induction tree* dapat dilakukan dengan *binary split* maupun *multiway split*. Jika hasil kondisi tes terbagi menjadi dua yaitu ($A < v$) atau ($A > v$), maka menggunakan *binary split*. Jika hasil kondisi tes terbagi menjadi beberapa bagian, maka menggunakan *multiway split*. *Induction tree* berdasarkan *continuous attribute* ditunjukkan Gambar 2.7



Gambar 2.7 Contoh Decision Tree

Random forest terdapat beberapa tree. Setiap tree merupakan independent bootstrap sample dari data training. Parameter yang biasanya digunakan di dalam metode Random Forest yaitu *n_estimators* sebagai jumlah dari tree di dalam forest, *Max_depth* adalah kedalaman maksimal dari tree. Jika

variable `max_depth` tidak inisiasi nilainya, node akan terus berkembang sampai semua *leaves* terdiri dari kurang dari variable `min_sample_split`. Sedangkan variable `random_state` digunakan untuk generator bilangan random. Bagaimana decision tree ini menangani permasalahan regresi. merupakan contoh data yang akan diselesaikan secara decision tree.

Algoritma dalam Random Forest adalah:

Untuk setiap node dilakukan :

1. Pilih sejumlah n variable dan lakukan random semua kemungkinan kombinasi variabel.
2. Temukan *split* yang terbaik pada variabel yang terpilih.
3. Nilai prediksi adalah rata rata dari semua trees yang diobservasi menggunakan decision tree regresi

Di dalam Tabel 2.5 diberikan contoh data untuk kasus regresi, dimana terdapat 4 variabel yang menentukan yaitu Outlook, Temp, Humidity dan Windy yang mempengaruhi lamanya waktu bermain (hour played)

Tabel 2.5 contoh data regresi

Outlook	Temp	Humidity	Windy	Hour Played
Rainy	Hot	High	FALSE	25
Rainy	Hot	High	TRUE	30
Overcast	Hot	High	FALSE	46
Sunny	Mild	High	FALSE	45
Sunny	Cool	Normal	FALSE	52
Sunny	Cool	Normal	TRUE	23
Overcast	Cool	Normal	TRUE	43
Rainy	Mild	High	FALSE	35
Rainy	Cool	Normal	FALSE	38
Sunny	Mild	Normal	FALSE	46
Rainy	Mild	Normal	TRUE	48
Overcast	Mild	High	TRUE	52
Overcast	Hot	Normal	FALSE	44
Sunny	Mild	High	TRUE	30

Pembangunan decision tree dimulai dengan menemukan cabang yang paling homogen atau root dari tree. Langkah yang dilakukan

1. Menghitung standard deviasi dari satu atribut yaitu target dengan persamaan 2.21

$$S = \sqrt{\frac{\sum (x - \mu)^2}{n}} = 9.32 \quad (2.21)$$

Sedangkan standard deviasi dari 2 atribut menggunakan persamaan

$$S(T, X) = \sum_{c \in X} P(c) S(c) \quad (2.22)$$

2. Datasets dibagi ke dalam atribut yang berbeda beda yang ditunjukkan di dalam Tabel 2.6. Kemudian Standard deviasi dari tiap cabang dihitung. Standard deviasi yang dihasilkan dikurangkan dengan standard deviasi sebelum proses split. Hasil ini dinamakan dengan Standard Deviasi Reduction atau SDR seperti yang ditunjukkan di dalam persamaan 2.22.

$$SDR(T, X) = S(T) - S(T, X) \quad (2.23)$$

Tabel 2.6 Pembagian dataset berdasarkan atribut

Overcast		Rainy		Sunny	
Hours Played	$x - \mu^2$	Hours Played	$x - \mu^2$	Hours Played	$x - \mu^2$
46	0.06	25	104	45	34
43	10.6	30	27	52	164
52	33.1	35	0.04	23	262
44	5.06	38	7.84	46	46

Setelah pembagian tersebut, dihasilkan nilai standard deviasi dari setiap atribut seperti yang ditunjukkan di dalam Tabel 2.7

Tabel 2.7 Standard deviasi dari data

Atribut	Standard Deviasi	Jumlah
Overcast	3.49	4
Rainy	7.78	5
Sunny	10.87	5

Standard Deviasi atribut Hours Played dan Outlook

$$S(\text{Hours}, \text{Outlook}) = (4/14) \cdot 3.49 + (5/14) \cdot 7.78 + (5/14) \cdot 10.87 = 7.66$$

Sedangkan Standard Deviasi Reductionnya

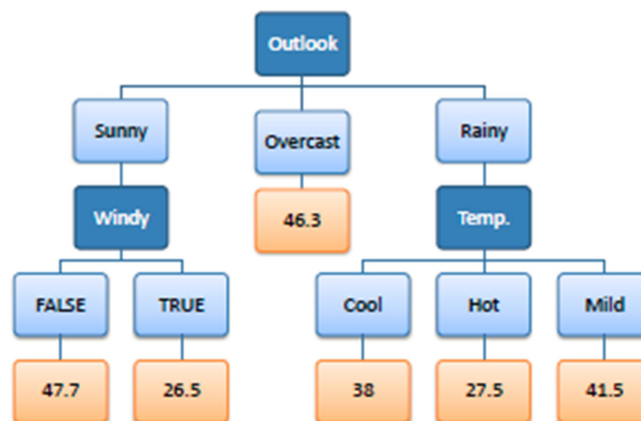
$$\begin{aligned} \text{SDR}(\text{Hours}, \text{Outlook}) &= S(\text{Hours}) - S(\text{Hours}, \text{Outlook}) \\ &= 9.32 - 7.66 = 1.66 \end{aligned}$$

$$\begin{aligned} \text{SDR}(\text{Hours}, \text{Temp}) &= S(\text{Hours}) - S(\text{Hours}, \text{Temp}) \\ &= 9.32 - 8.84 = 0.48 \end{aligned}$$

$$\begin{aligned} \text{SDR}(\text{Hours}, \text{Humidity}) &= S(\text{Hours}) - S(\text{Hours}, \text{humidity}) \\ &= 9.32 - 9.05 = 0.27 \end{aligned}$$

$$\begin{aligned} \text{SDR}(\text{Hours}, \text{Windy}) &= S(\text{Hours}) - S(\text{Hours}, \text{Windy}) \\ &= 9.32 - 9.04 = 0.28 \end{aligned}$$

3. Atribut yang mempunyai Standard Deviasi Reduction paling besar dipilih sebagai Decision Node atau Root. Hasil dari proses decision tree ini ditunjukkan oleh Gambar 2.8.



Gambar 2.8 Hasil dari Decision Tree regresi

2.9 Optimisasi Hyperparameter

Yang membedakan antara *hyperparameter* dan parameter adalah nilai *hyperparameter* tidak ditentukan dari pelatihan sebuah model, tapi ditentukan sebelum proses training dilakukan. Penentuan suatu parameter yang digunakan dalam suatu model bukan sesuatu yang mudah. Jika kombinasi variable dicari

secara manual, maka akan membutuhkan waktu yang cukup lama dan sulit menemukan kombinasi yang tepat untuk memperoleh model dengan error yang terkecil. Di dalam penelitian ini digunakan dua metode optimisasi yaitu Grid Search dan Random Search.

2.9.1 Metode optimisasi hyperparameter Grid Search

Grid search merupakan salah satu metode konvensional dalam menentukan parameter yang optimal untuk sebuah model. Langkah pertama yang dilakukan yaitu menentukan range atau jangkauan parameter yang akan digunakan, kemudian akan dilakukan pengecekan parameter mana yang menghasilkan akurasi yang lebih baik. Algoritma Grid Search mencari pasangan parameter yang menghasilkan kombinasi dan hasil yang terbaik. Hasil dari Grid Search digunakan di dalam model prediksi. Terdapat kelemahan dari penggunaan Grid Search, proses pencarian parameter yang optimal dengan metode Grid Search membutuhkan waktu yang lama. Oleh karena itu metode ini sering disebut metode cost expensive.

Contoh Cara memilih hyperparameter misalkan pemilihan parameter ke-1, p1 dan parameter ke-2, p2 maka kombinasi kedua parameter dengan cara Grid Search ditampilkan dalam tabel 2.8 dari kombinasi tersebut bisa dievaluasi (bisa menggunakan evaluasi error mean squared error / mse) untuk memilih kombinasi manakah yang terbaik.

$$p1 = [1,10,100]$$

$$p2 = [1e3, 1e4, 1e5]$$

Tabel 2.8 Kombinasi parameter menggunakan metode Grid Search

	P1		
P2	(1,1e3)	(10,1e3)	(100,1e3)
	(1,1e4)	(10,1e4)	(100,1e4)
	(1,1e5)	(10,1e5)	(100,1e5)

2.9.2 Metode Optimisasi Hyperparameter Random Search

Metode *Random Search*, mencari parameter secara random atau acak. Pencarian parameter yang optimal menggunakan *Random Search* lebih cepat dibandingkan menggunakan *Grid Search* tetapi akurasi parameter *Grid Search* biasanya masih lebih baik dibandingkan *Random search*. Di dalam *Random Search*, semua kemungkinan parameter tidak diujikan seperti halnya yang dilakukan oleh metode *Grid Search*.

Algoritma Random Search

1. Inisialisasi parameter model. Misalkan jika mencari parameter model SVR kernel RBF, maka yang akan cari parameternya adalah parameter C, gamma dan epsilon (jika menggunakan SVR epsilon). Parameter inisiasi yang bisa digunakan adalah $C = 1$, $\gamma = 0.1$ dan $\epsilon = 0.1$ kemudian inisialisasi juga parameter iterasi yang digunakan (misalkan iterasi = 100) dan range parameter yang akan dicari (misalkan $C = \{1,2,3,4,5\}$)
2. Sampai kondisi iterasi belum tercapai atau kondisi berhenti belum ditemukan maka proses berikut dilakukan :
 - a. Mencari error terkecil dengan menggunakan parameter tersebut
 - b. Jika $f(y) < f(x)$ kemudian setting parameter x dengan y, artinya parameter yang ditemukan tersebut mempunyai error terkecil daripada model parameter sebelumnya.
3. Simpan x sampai ditemukan parameter terbaik lagi, jika kondisi berhenti belum terpenuhi, iterasi ditambahkan dengan 1 kemudian ulangi langkah ke-2

Pembangunan prediksi IHK dengan metode SVR, menggunakan parameter yang perlu diinisialisasi nilainya terlebih dahulu. Kemudian range parameter juga ditentukan.

2.10 Tool komputasi mesin pembelajaran

Tool komputasi yang digunakan didalam penelitian ini adalah python. Pemrograman *Python* dibuat oleh Guido van Rossum pada tahun 1991. *Python* merupakan bahasa pemrograman yang mempunyai *library* yang sangat banyak sehingga bisa menyelesaikan banyak masalah. Selain *library* yang banyak, *python* mempunyai management memory yang otomatis dan didukung oleh banyak sistem operasi. Pemrograman ini juga tidak berbayar sehingga dapat dikembangkan dan tidak ada batasan dalam penyalinan dan pendistribusian. *Python* mempunyai beberapa layanan yang disediakan lengkap dengan source codenya, debugger dan profile, interface, fungsi sistem, GUI. Python tersedia untuk berbagai sistem Operasi, Windows, Linux tetap bisa dijalankan. Kemudahan inilah yang membuat *python* banyak digunakan peneliti, data analitik dan developer untuk mendukung aktifitasnya. Data analitik biasanya menggunakan python sebagai mesin pembelajaran. Beberapa dari *Library Python* yang sering digunakan di dalam data analitik adalah :

- NumPy mendukung komputasi scientific multidimensional misalnya array dan matrix dan fungsi matematika high-level
- SciPy mendukung perhitungan scientific misalkan perhitungan integral, turunan
- Scikit-learn mendukung implementasi algoritma mesin pembelajaran yang mendukung data analitik
- Matplotlib digunakan untuk menampilkan grafik atau 2D plotting
- BeautifulSoup Library ini bisa digunakan untuk teknik web scrapping dan crawling data twitter. Library ini membantu parsing xml dan html
- Nltk Natural Language Toolkit - jika ingin melakukan manipulasi string, library ini sangat membantu.

Jupyter Notebook

Jupyter merupakan antar muka web dengan menggunakan *shell python*. Jupyter notebook ini memungkinkan penggunanya melakukan proses edititng dan running document melalui web browser. Jupyter notebook ini merupakan aplikasi

server client. Selain bisa diakses secara remote, aplikasi ini bisa dapat diakses di desktop local dan tanpa membutuhkan akses internet. *Dashboard dari notebook* merupakan tampilan yang pertama kali ditunjukkan pada saat pengguna masuk pertama kali.

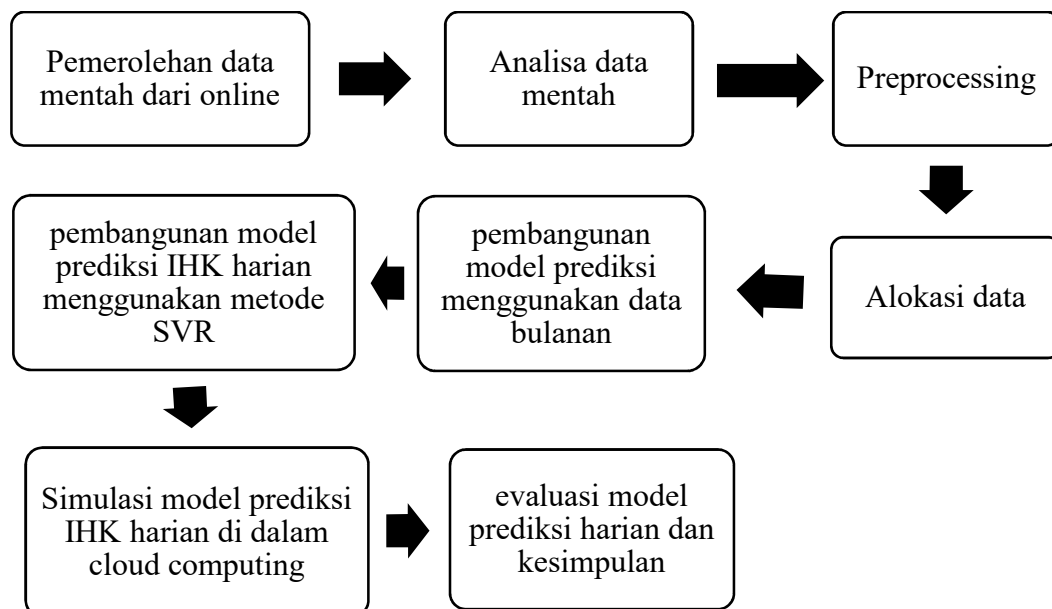
BAB 3

METODOLOGI PENELITIAN

Di dalam bab ini akan dijelaskan metode dan cara kerja yang akan digunakan di dalam penelitian, sehingga dapat memberikan gambaran bagaimana membangun model prediksi nilai IHK harian menggunakan Support Vector Regression dan melakukan simulasi prediksi jika dijalankan di dalam lingkungan Cloud Computing.

3.1 Tahapan Penelitian

Tahapan penelitian yang digunakan dalam penelitian ini terdiri dari delapan fase **utama** yaitu pemerolehan data mentah dari online, analisa data mentah, preprocessing, alokasi data, pembangunan model prediksi menggunakan data bulanan, pembangunan model prediksi IHK harian menggunakan metode SVR, simulasi model prediksi IHK harian di dalam cloud computing dan fase terakhir adalah proses evaluasi model prediksi harian dan kesimpulan. Tahapan penelitian yang dilakukan pada penelitian ini ditunjukkan di dalam Gambar 3.1.



Gambar 3.1 Diagram tahapan penelitian

3.2 Pemerolehan data mentah

Dalam penelitian ini, variable indikator inflasi yaitu IHK akan diprediksi. Data yang digunakan dalam penelitian ini adalah data mentah dari online data. Data IHK tersebut diperoleh dari data online BPS [9] yang dipublikasi dengan format dokumen pdf. Data IHK inilah yang akan dijadikan variable target. Sedangkan untuk variable masukan, menggunakan data harga konsumen komoditas bahan pokok dan data nilai tukar rupiah terhadap dollar yang dipublikasi online di website. Kedua macam data tersebut dapat diambil dengan teknik *web scrapping* dan *text mining* file (dalam format file pdf). Teknik *web scrapping* ini memungkinkan data diambil secara online di sebuah halaman web kemudian akan dimasukkan kedalam file dengan format csv untuk kemudian dilakukan proses analisa dan pengolahan data. Di dalam *text mining*, data yang diinginkan bisa diperoleh dengan melakukan teknik *text mining* file dokumen.

Data harga pokok online diperoleh dari website Sistem Informasi Ketersediaan dan Perkembangan Harga Bahan Pokok di Jawa Timur (Siskaperbapo.com) yang dikelola Dinas Perindustrian dan Perdagangan Jawa Timur [10]. Data di dalam web Siskaperbapo berisi data harga konsumen, harga produsen dan data ketersediaan yang diperbarui setiap hari. Data Siskaperbapo yang digunakan di dalam penelitian ini adalah data harga konsumen. Data harga konsumen tersebut berasal dari 38 kota/kabupaten di Jawa timur dari tahun 2011. Data di tahun 2011 masih berupa data percobaan, jadi didalam penelitian ini hanya menggunakan data dari bulan Januari tahun 2012 sampai bulan Desember tahun 2016. Dari 38 data kota di Jawa timur, penelitian ini hanya menggunakan 8 data kota/kabupaten yang mewakili proses sampling IHK oleh BPS. Data 8 kota itu meliputi Kota Surabaya, Kota Madiun, Kota Kediri, Kabupaten Sumenep, Kabupaten Jember, Kota Probolinggo, Kabupaten Banyuwangi dan Kota Malang. Data harga konsumen terdiri dari 28 jenis bahan pokok dan terbagi menjadi 68 item bahan pokok, ditunjukkan di dalam Tabel 3.1, yaitu tabel contoh Data Harga Konsumen dari Siskaperbapo.com.

Tabel 3.1 Contoh Data Harga Konsumen Dari Siskaperbapo.com

No	Nama Bahan Pokok	Harga Kemarin	Harga Sekarang
1	Beras		
	- Bengawan	10.289	10.339
	- Mentik	10.081	10.097
	- Ir 64	8.778	8.778
2	Gula Pasir		
	- Gula Pasir Dalam Negri	12.327	12.324
3	Minyak Goreng		
	- Bimoli Botol / Kemasan (Sps) 620ml	11.819	12.024
	- Bimoli Botol/Kemasan (Sps) 2 Liter	26.44	26.452
	- Tanpa Merk / Minyak Curah	11.537	11.517
4	Daging		
	- Daging Sapi Murni	106.943	106.53
	- Daging Ayam Broiler	29.682	29.73
	- Daging Ayam Kampung	56.818	57.063
5	Telur Ayam		
	- Telur Ayam Ras / Petelur	18.119	18.224
	- Telur Ayam Kampung	35.949	35.826
6	Susu		
	Kental Manis		
	- Susu Kental Manis Merk Bendera	9.938	9.969
	- Susu Kental Manis Merk Indomilk	9.698	9.698
	Susu Bubuk		
	- Susu Bubuk Merk Bendera (Instant)	38.132	38.259
	- Susu Bubuk Merk Indomilk (Instant)	35.665	35.762
7	- Jagung Pipilan Kering	6.081	6.122
8	- Garam Beryodium	-	-
	- Bata	735	734
	- Halus	4.466	4.516
9	Tepung Terigu		
	- Segitiga Biru (Kw Medium)	7.689	7.642
10	Kacang Kedelai		
	- Eks Impor	8.392	8.386
	- Lokal	8.357	8.334
11	Mie Instant		
	- Indomie Rasa Kari Ayam	2.217	2.231
12	Cabe		
	- Keriting	19.58	19.336
	- Biasa	17.986	17.933
	- Cabe Rawit	30.06	29.494
13	- Bawang Merah	25.301	25.146
	- Bawang Putih	37.137	36.981
14	- Ikan Asin Teri	61.105	61.865
15	- Kacang Hijau	18.501	18.425
16	- Kacang Tanah	24.142	23.953
17	- KETELA POHON	3.248	3.333

Tabel 3.2 Contoh Data Harga Konsumen Dari Siskaperbapo.com (Lanjutan)

No	Nama Bahan Pokok	Harga Kemarin	Harga Sekarang
18	Sayur Mayur		
	- Kol/Kubis	6.211	6.181
	- Kentang	13.702	13.691
	- Tomat	7.601	7.419
	- Wortel	11.736	11.705
	- Buncis	8.114	8.429
19	Semen		
	- Semen Gresik	49.606	49.681
	- Semen Tiga Roda	47.777	47.955
	- Semen Holcim	46.62	46.791
	- Semen Padang	46.104	46.104
	- Semen Tonasa	46.019	45.944
	- Semen Bosowa	46.244	46.313
20	Ikan Segar		
	- Ikan Bandeng	25.052	25.24
	- Ikan Kembung	26.548	26.539
	- Ikan Tuna	32.455	32.518
	- Ikan Tongkol	25.914	25.944
	- Ikan Cakalang	26.935	26.994
22	- Kayu Balok Meranti (4 X 10)	93.021	93.132
23	- PAPAN MERANTI (4m X 3cm X 20mm)	98.765	98.588
24	- Triplek (6mm)	74.83	74.82
25	Besi Beton (Sni Murni)		
	- Besi Beton 6 Mm (12/9m)	21.252	21.224
	- Besi Beton 8 Mm (12/9m)	32.505	32.959
	- Besi Beton 10 Mm (12/9m)	50.949	50.848
	- Besi Beton 12 Mm (12/9m)	70.75	70.798
26	Paku		
	- Paku Ukuran 10cm	12.852	12.879
	- Paku Ukuran 2 Cm	14.556	14.635
	- Paku Ukuran 3cm	13.986	14.096
	- Paku Ukuran 4cm	13.618	13.695
	- Paku Ukuran 5cm	13.262	13.288
	- Paku Ukuran 7cm	13.032	13.049
27	- Gas Elpigi 3 Kg	17.495	17.49
28	Pupuk		
	- Pupuk Kel Non Subsidi	5.717	5.676
	- Pupuk Npk Non Subsidi	6.83	6.952
	- Pupuk Sp 35 Non Subsidi	3.496	3.467
	- Pupuk Urea Non Subsidi	3.597	3.522
	- Pupuk Za Non Subsidi	3.046	3.018

3.3 Analisa data mentah yang akan digunakan dalam prediksi IHK harian

Proses analisa data mentah ini perlu dilakukan untuk memilih variable harga konsumen atau sumber data yang layak digunakan didalam penelitian. Yang dimaksudkan data yang layak yang dimaksudkan adalah data yang ideal tanpa banyak kekosongan nilai dalam setiap variabel harga konsumen atau setiap bahan pokok.

Salah satu analisa yang dilakukan, variabel harga bahan konsumen yang disurvey pada tahun 2012 tidak sama dengan variable di tahun 2016. Pada tahun 2012, hanya dilakukan survey 41 harga bahan pokok atau ditunjukkan di Tabel 3.1 no. 1 sampai dengan no 19, sedangkan pada tahun 2016 telah disurvey 28 jenis bahan atau 68 item. Karena perbedaan bahan pokok yang disurveykan, banyak nilai yang hilang dalam 27 variable lainnya di tahun 2012 (di dalam Tabel 3.1 variabel yang banyak tidak ada nilainya adalah harga bahan pokok no. 20 s.d 28).

Dari 68 variable harga bahan pokok yang terlihat di Tabel 3.1, hanya 30 bahan pokok yang dipilih untuk digunakan sebagai variabel masukan. Hal ini dikarenakan 38 item bahan pokok lainnya mempunyai banyak nilai yang hilang (*missing value*).

Pada Table 3.2, tabel Ringkasan data nilai yang hilang di delapan kota Jawa Timur, data tersebut menunjukkan banyaknya data yang hilang dalam delapan kota di Jawa Timur. Nilai yang hilang inilah yang dapat mempengaruhi akurasi prediksi. Untuk mengatasinya, ada beberapa cara yang bisa dilakukan, salah satunya adalah dengan cara menghilangkan nilai yang hilang itu dari dataset. Tetapi ada yang perlu diperhatikan, data yang dihilangkan tersebut seharusnya tidak membuat akurasi prediksi IHK menjadi berkurang. Jika variable tersebut dihilangkan, bisa jadi variable tersebut adalah variable yang penting bagi model prediksi, sehingga akurasi prediksi bisa menurun.

Warna hijau di dalam Tabel 3.2, yaitu dalam variable harga bahan pokok Minyak Bimoli 2lt dan Beras Bengawan, variable tersebut tidak dihilangkan, tetapi akan dicari nilai untuk mengisi kekosongan data tersebut. Variable ini tidak dihilangkan karena nilai kosong yang terbesar hanya pada satu kota, Minyak

Bimoli 2lt nilai kosong terbesar di kota Malang dan Beras Bengawan nilai kosong terbesar di kabupaten Banyuwangi.

Tabel 3.3 Ringkasan data nilai yang hilang di 8 kota Jawa Timur

No	Bahan Pokok	Sby	Mlg	Kdr	Pbg	Mdn	Smnp	Jbr	Bwi
1	BERAS								
	Bengawan	0	0	11	39	44	43	16	1505
	Mentik	31	31	11	1827	44	1365	106	1506
	IR 64	0	0	11	30	44	19	16	2
2	GULA PASIR								
	Gula Pasir Dalam Negri	0	0	11	30	44	16	16	2
3	MINYAK GORENG								
	Bimoli Spesial 620 ml	1637	0	203	1618	131	546	688	7
	Bimoli botol 2 liter	0	1512	11	30	45	16	16	3
	Tanpa Merk	0	0	11	30	46	16	16	2
4	DAGING								
	Daging Sapi Murni	5	5	11	30	44	16	16	2
	Daging Ayam Broiler	0	0	11	30	44	16	16	2
	Daging Ayam Kampung	0	0	11	40	46	16	16	2
5	TELUR AYAM								
	Telur Ayam Ras/ Petelur	0	0	11	30	44	16	16	2
	Telur Ayam Kampung	0	325	11	42	45	16	16	2
6	SUSU								
	Susu Kental Manis								
	Susu Kental Manis Merk Bendera	0	0	11	30	45	16	16	2
	Susu Kental Manis Merk Indomilk	0	0	11	30	45	16	16	2
	Susu Bubuk								
	Susu Bubuk Merk Bendera	0	0	11	161	45	232	17	702
	Susu Bubuk Merk Indomilk	0	0	11	161	45	250	127	725
7	JAGUNG PIPILAN KERING	0	0	11	30	45	16	16	2
8	GARAM BERYODIUM								
	Bata	0	0	11	30	45	16	16	36
	Halus	0	0	11	30	45	16	17	2
9	TEPUNG TERIGU								
	Segitiga Biru	0	0	11	30	45	16	16	2
10	KACANG KEDELAI								
	Eks Impor	0	0	11	716	87	20	181	2
	Lokal	95	328	11	831	44	19	419	3
11	Indomie Rasa Kari Ayam	0	0	11	30	44	16	16	2

Tabel 3.4 Ringkasan data nilai yang hilang di 8 kota Jawa Timur (Lanjutan)

No	Bahan Pokok	Sby	Mlg	Kdr	Pbg	Mdn	Smnp	Jbr	Bwi
12	CABE								
	- Keriting	460	16	11	1748	43	453	938	1749
	- Biasa	0	0	11	30	43	16	16	2
	- Cabe Rawit	0	0	11	30	43	112	18	2
13	- Bawang Merah	0	0	11	30	46	16	16	2
	- Bawang Putih	0	0	11	30	48	16	16	2
14	Ikan Asin Teri	0	61	11	30	49	21	16	2
15	Kacang Hijau	0	0	11	30	47	16	16	2
16	Kacang Tanah	0	0	11	30	47	16	16	2
17	Ketela Pohon	0	0	11	62	48	16	16	2
18	SAYUR MAYUR								
	- Kol/Kubis	0	0	11	48	48	16	16	2
	- Kentang	0	0	11	48	49	16	16	2
	- Tomat	0	0	11	48	49	16	16	2
	- Wortel	0	0	11	48	49	16	16	2
	- Buncis	0	0	11	80	48	48	16	2

Keterangan kode :

Sby=Surabaya, Mlg=Malang, Kdr=Kediri, Pbg=Probolinggo, Mdn=Madiun,
Smnp=Sumenep, Jbr=Jember, Bwi=Banyuwangi

Ringkasan **variable** harga bahan pokok yang hilang, ditunjukkan dalam data di Tabel 3.4 yaitu tabel ringkasan data yang hilang tiap variable. Dalam Tabel 3.4 nilai harga bahan pokok yang hilang setiap variabel, ditunjukkan dengan warna merah yaitu pada bahan pokok Telur Ayam Kampung, Susu Bubuk Merk Bendera, Cabe Keriting, kacang kedelai Eks Impor, kacang kedelai Lokal, Susu Bubuk Merk Bendera, Susu Bubuk Merk Indomilk. Sedangkan warna hijau adalah variabel yang hilang tetapi merujuk dari Tabel 3.3, nilai yang hilang pada variabel Beras Mentik dan Bimoli Spesial 2lt ini hanya besar di salah satu kota, sehingga masih memungkinkan, nilai yang hilang ini diatasi dengan mengisinya atau imputed missing value.

Tabel 3.5 Ringkasan data yang hilang tiap variable

No	Variabel Bahan Pokok	Nilai hilang	No	Variabel Bahan Pokok	Nilai hilang
	Beras			Garam Beryodium	
1	- Bengawan	1658	18	- Bata	154
2	- Mentik	4921	19	- Halus	121
3	- IR 64	122		Tepung Terigu	
	Gula Pasir		20	Segitiga Biru	120
4	- Gula Pasir Dalam Negri	119		Kacang Kedelai	
	Minyak Goreng		21	- Eks Impor	1017
5	- Bimoli spesial 620ml	4830	22	- Lokal	1750
6	- Bimoli spesial 2 liter	1633		Mie Instant	
7	- Tanpa Merk	121	23	- Indomie Kari Ayam	119
	Daging			Cabe	
8	- Daging Sapi Murni	129	24	- Keriting	5418
9	- Daging Ayam Broiler	119	26	- Biasa	118
10	- Daging Ayam Kampung	131	27	- Cabe Rawit	216
	Telur ayam		28	- Bawang Merah	121
11	- Telur Ayam Ras / Petelur	119	29	- Bawang Putih	123
12	- Telur Ayam Kampung	457	30	- Ikan Asin Teri	190
13	- Susu Kental Manis Merk Bendera	120	31	- Kacang Hijau	122
14	- Susu Kental Manis Merk Indomilk	120	32	- Kacang Tanah	122
	Susu		33	- Ketela Pohon	155
	Kental Manis			Sayur Mayur	
13	- Susu Kental Manis Merk Bendera	120	34	- Kol/Kubis	141
14	- Susu Kental Manis Merk Indomilk	120	35	- Kentang	142
	Susu Bubuk		36	- Tomat	142
15	- Susu Bubuk Merk Bendera	1168	37	- Wortel	142
16	- Susu Bubuk Merk Indomilk	1319	38	- Buncis	205
17	Jagung Pipilan Kering	120			

Variable bahan pokok dengan warna merah tersebut akan dihilangkan dan tidak digunakan dalam dataset. Setelah proses memilih variabel yang

digunakan dalam prediksi maka pada Table 3.5, tabel variable yang digunakan di dalam model prediksi IHK, menunjukkan variable yang dipilih dan digunakan untuk nilai masukan X di dalam model prediksi IHK.

Tabel 3.6 Variable yang digunakan di dalam model prediksi IHK

No	Nama variabel	No	Nama variabel
x1	Beras Bengawan	x16	Indomie rasa kari ayam
x2	Beras IR 64	x17	Cabai biasa
x3	Gula Pasir Dalam Negri	x18	Cabai rawit
x4	Minyak Bimoli 2 Liter	x19	Bawang merah
x5	Minyak Tanpa Merk	x20	Bawang putih
x6	Daging Sapi Murni	x21	Ikan asin teri
x7	Daging Ayam Broiler	x22	Kacang hijau
x8	Daging Ayam Kampung	x23	Kacang tanah
x9	Telur Ayam Ras / Petelur	x24	Ketela pohon
x10	Susu Kental Manis Merk Bendera	x25	Kol/kubis
x11	Susu Kental Manis Merk Indomilk	x26	Kentang
x12	Jagung pipilan kering	x27	Tomat
x13	Garam bata	x28	Wortel
x14	Garam halus	x29	Buncis
x15	Tepung segitiga biru	x30	Kurs dollar

Dari penjelasan diatas, disimpulkan bahwa di dalam penelitian ini menggunakan :

variabel target $y = \text{IHK bulanan}$

dan

variabel masukan $X = x_1, x_2, x_3, x_4, x_5, x_6, \dots, x_{30}$

Data masukan atau variabel X dibagi menjadi data bulanan dan data harian. Di dalam data bulanan variabel masukan X, menggunakan rata-rata harga konsumen tiap bulan dan variabel target, y menggunakan nilai IHK bulanan yang bersumber dari BPS. Dalam data harian, variabel masukan X adalah data harian harga bahan pokok dan nilai tukar rupiah terhadap dollar (kurs). Variable target y, menggunakan IHK harian dan nilai IHK harian inilah yang diestimasi nilainya menggunakan model bulanan.

Dalam Tabel 3.6 menunjukkan contoh dataset bulanan dari Kota Surabaya dari bulan Januari tahun 2012 sampai dengan bulan Oktober tahun 2013,

untuk variabel X1, X2, X3, X4, X27, X28, X29, X30. untuk data bulanan lengkap dilampirkan dalam buku ini. Dataset inilah yang akan digunakan di dalam prediksi IHK dengan menggunakan metode Support Vector Regression

Tabel 3.7 Contoh data bulanan Kota Surabaya dari Tahun 2012 - 2016

No	Bulan	X1	X2	X3	X4	X27	X28	X29	X30	IHK
1	Jan-12	9180	7652	9641	23997	10557	6338	5488	9160	98.2
2	Feb-12	9206	7703	9852	24000	6550	5756	6177	9071	98.55
3	Mar-12	9207	7669	10733	25019	6007	5435	6528	9211	98.77
4	Apr-12	9226	7615	10975	26993	6300	4775	5249	9222	99.01
5	May-12	9094	7660	11600	27000	4601	4520	4851	9337	98.85
6	Jun-12	9176	7725	12273	26973	5206	4611	5197	9498	99.34
7	Jul-12	9261	7747	12174	26871	4825	4253	5271	9504	100.06
8	Aug-12	9253	7740	11328	26400	4966	5239	5672	9547	101.19
9	Sep-12	9240	7683	11183	26400	5410	5808	5896	9614	101.04
10	Oct-12	9258	7715	11037	26400	5046	7018	6918	9645	101.03
11	Nov-12	9285	7776	11359	26400	4777	8982	7261	9676	101.2
12	Dec-12	9342	7808	11161	26400	4871	9474	6849	9694	101.54
13	Jan-13	9497	7800	10985	26400	6212	9700	7010	9735	102.95
14	Feb-13	9600	7800	11059	26400	9100	9593	6893	9735	103.73
15	Mar-13	9634	7800	11081	26400	9290	8023	6403	9758	104.73
16	Apr-13	9626	7808	11109	26400	7547	6467	6557	9773	104.34
17	May-13	9638	7920	11431	26400	6710	7071	6561	9810	103.6
18	Jun-13	9637	7931	11310	26579	7590	8443	6647	9931	104.41
19	Jul-13	9763	8014	11341	26170	10313	10550	7210	10124	108.17
20	Aug-13	9603	7872	10997	23839	11155	9950	6626	10625	109.17
21	Sep-13	9635	7914	10990	23540	6237	6884	5393	11403	108.35
22	Oct-13	9868	7841	10780	23423	4966	7039	6081	11424	108.56

3.4 Tahapan Pengolahan awal data atau Preprocessing

Tahap ini adalah tahap persiapan data agar dapat diolah lebih lanjut menggunakan metode data mining. Hal ini dilakukan untuk mendapatkan hasil prediksi IHK yang akurasinya baik.

3.4.1 Mengubah nilai IHK ke dalam tahun dasar yang sama

BPS tidak hanya mensurvey harga barang secara berkala, tetapi juga melakukan survei biaya hidup untuk membuat Diagram Timbang Nilai Konsumsi. Data tersebut menunjukkan pola konsumsi masyarakat. Diagram timbang tersebut digunakan untuk menghitung IHK masing-masing kota. IHK menggunakan skala 100 dan dilakukan pembaruan daftar timbang setiap lima tahun untuk menyesuaikan daftar timbang dengan kondisi dan kebutuhan yang terjadi. Dalam penelitian ini, data IHK bulanan di bulan Januari tahun 2012 s.d bulan Desember tahun 2016 menggunakan dua tahun dasar yaitu 2007 dan 2012. Untuk membuat model peramalan yang baik, model sebaiknya hanya menggunakan satu tahun dasar. Metode menyamakan tahun dasar yang digunakan adalah persamaan 3.1, persamaan ini merupakan turunan dari persamaan laju inflasi (3.1).

$$IHK_{n-1} = \frac{IHK_n}{100 + inflation_{rate\ n}} \quad (3.2)$$

dimana :

IHK_n : Indeks Harga Konsumen di bulan n

$IHK_{(n-1)}$: Indeks Harga Konsumen di bulan n-1

$Inflasi_n$: Nilai Inflasi di bulan n

3.4.2 Mengatasi permasalahan nilai yang hilang

Pembangunan model prediksi yang baik bergantung dengan data yang dilatihkan. Jika terdapat banyak nilai yang hilang dalam dataset, maka akan mempengaruhi akurasi prediksi model. Variable yang mengandung banyak nilai yang hilang tidak digunakan didalam penelitian ini. Sedangkan untuk mengatasi nilai hilang yang tidak banyak, dilakukan *imputed value* atau mengisi nilai yang hilang. Untuk mengisi nilai yang hilang, nilai mean, median dan nilai data

sebelumnya dapat digunakan, tergantung dari posisi data yang hilang. Jika terletak di dalam range dua tanggal maka dapat digunakan nilai median, jika nilai yang hilang tidak berada di dalam range dua tanggal maka digunakan mean data yang sejenis.

3.4.3 Pengubahan skala data

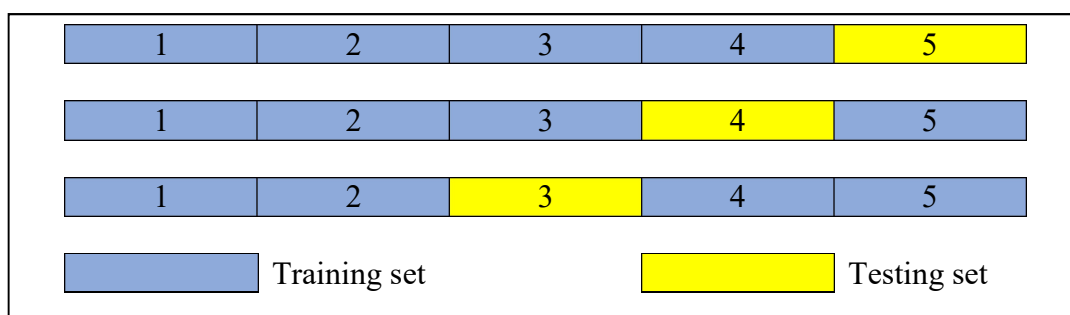
Metode ini dilakukan agar nilai data berada pada range yang sama dan mendapatkan hasil akurasi yang lebih akurat. Di dalam penelitian ini, data variable masukan, X diubah menjadi skala (0,1). Pengubahan skala data tersebut menggunakan persamaan (2.1)

3.5 Alokasi data

Di dalam pembelajaran mesin prediksi, diperlukan dataset untuk dilatihkan sehingga membentuk sebuah model prediksi yang diinginkan. Jika data training yang dipelajari atau dilatih sama dengan yang diujikan, itu sama halnya dengan mengulang label atau sampel yang telah dilihat model. Model prediksi tersebut biasanya mempunyai hasil prediksi sempurna namun gagal memprediksi data lainnya, kondisi ini sering disebut *overfitting*. Untuk menghindari kejadian itu, maka dibutuhkan data pengujian yang berbeda dengan data latih. Di dalam penelitian ini, data dibagi menjadi dua bagian, yaitu data training dan data testing. Di dalam dataset prediksi IHK bulanan dan harian, alokasi data training sebesar 70% sedangkan 30% sebagai data testing. Untuk menunjukkan perbandingan waktu komputasi dari penggunaan metode, pembagian data lebih bervariasi. Data dibagi menjadi tiga set yaitu set pertama data training sebesar 90% dan 10% data testing, set kedua sebesar 80% dan 20% dan set ketiga 70% training dan 30% data testing.

Dalam penelitian ini juga menggunakan metode *cross-validation k-fold*. Di dalam metode ini, dataset dibagi menjadi subset data training dan data testing kemudian dicari kombinasi yang terbaik. Dilakukan silang data satu subset dijadikan data testing dan subset sisanya dijadikan data training. Setiap kombinasi subset diujikan sehingga menghasilkan score. Score yang digunakan disini adalah R^2 atau koefisien determinan. Kombinasi yang terbaik antara alokasi data

training dan data testing, akan digunakan di dalam model prediksi. Indikator kombinasi subset data terbaik diperoleh jika model mempunyai R^2 mendekati nilai satu. *Cross-validation* yang digunakan dalam penelitian ini sama dengan 5-fold seperti yang ditunjukkan dalam Gambar 3.3, dimana warna kuning mewakili testing set dan warna biru artinya training set. Dalam percobaan pertama, subset 5 menjadi testing set dan subset 1,2,3 dan 4 menjadi training set. Kombinasi pertama di evaluasi apakah nilai R^2 nya mendekati nilai satu, Kemudian diujikan lagi pada kombinasi kedua dimana subset menjadi testing data dan subset 1,2,3 dan 5 menjadi training set. Begitu seterusnya sampai semua kombinasi telah dicobakan.



Gambar 3.2 Cross Validation dengan 5-fold

3.6 Pembangunan Model Prediksi IHK

3.6.1 Optimisasi Hyperparameter

Di dalam tahapan ini, akan ditentukan parameter terbaik model prediksi dan waktu komputasi yang dibutuhkan. Di dalam menentukan parameter, selain penentuan secara manual, dengan mencoba kombinasi setiap parameter sampai ditemukan koefisien determinan mendekat nilai satu atau MSE terkecil atau mendekati nilai nol, ada beberapa metode optimisasi parameter yang bisa digunakan. Di dalam penelitian ini menggunakan Grid Search dan Random Search. Akan dibuktikan, apakah metode Grid Search mempunyai waktu yang lebih lama dibandingkan dengan Random Search dalam kasus prediksi IHK. Jika penentuan parameter dilakukan secara manual, misalkan pada SVR dengan kernel Linear, parameter yang perlu ditentukan adalah C, gamma, dan epsilon. Maka harus mencari nilai tiap parameter dan kombinasinya sehingga menghasilkan performansi yang baik.

Pada tahap optimisasi hyperparameter akan ditentukan parameter mana yang paling sesuai untuk model prediksi IHK. Sebelum melakukan proses pemilihan parameter dan *fitting* (proses training data), dilakukan inisiasi parameter dan inisiasi range parameter. Range parameter disini untuk memberikan alternatif pilihan parameter yang akan diujikan. Setiap nilai di dalam range parameter ini akan diuji untuk mencari range dengan nilai evaluasi terbaik. Kombinasi terbaik diperoleh jika nilai evaluasi koefisien determinan yang dihasilkan bernilai mendekati nilai satu. Hyperparameter yang dioptimasi dan range parameter yang digunakan adalah :

1. Kernel Ridge Regression
 - $\text{Gamma} = \{0.01, 0.1, 1, 10, 100\}$
 - $\text{Alpha} = \{1, 0.1, 0.01, 0.001\}$
2. Support Vector Regression
 - a. Kernel Linear dan Radial basis function (rbf) mempunyai inisiasi parameter yang sama yaitu
 - $C = \{1, 2, 3, 4, 5, \dots, 99\}$, 99 nilai C
 - $\text{Gamma} = \{0.1, 1.1, 2.1\}$
 - $\epsilon (\text{epsilon}) = \{0.1, 1.1, 2.1\}$
 - b. Kernel Polynomial
 - $C = \{1, 2, 3, 4, 5, \dots, 99\}$
 - $\text{Gamma} = \{0.1, 1.1, 2.1\}$
 - $\epsilon (\text{epsilon}) = \{0.1, 1.1, 2.1\}$
 - $\text{degree} = \{1, 2, 3\}$
3. Random Forest Regressor
 - $\text{Max_depth} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
 - $\text{Random_state} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
 - $n_estimators = \{1, 2, 3, 4, 5, \dots, 19, 20\}$, 20 nilai n_estimator

Penerapan metode Grid Search pada hyperparameter Kernel Ridge Regression. Dimana di range di atas ada 5 nilai variabel $\text{Gamma} = \{0.01, 0.1, 1, 10, 100\}$ dan 4 nilai variabel $\text{Alpha} = \{1, 0.1, 0.01, 0.001\}$. Jika dibuatkan grid

parameternya akan terlihat seperti tabel 3.8, Penerapan Grid Search pada Kernel Ridge Regression

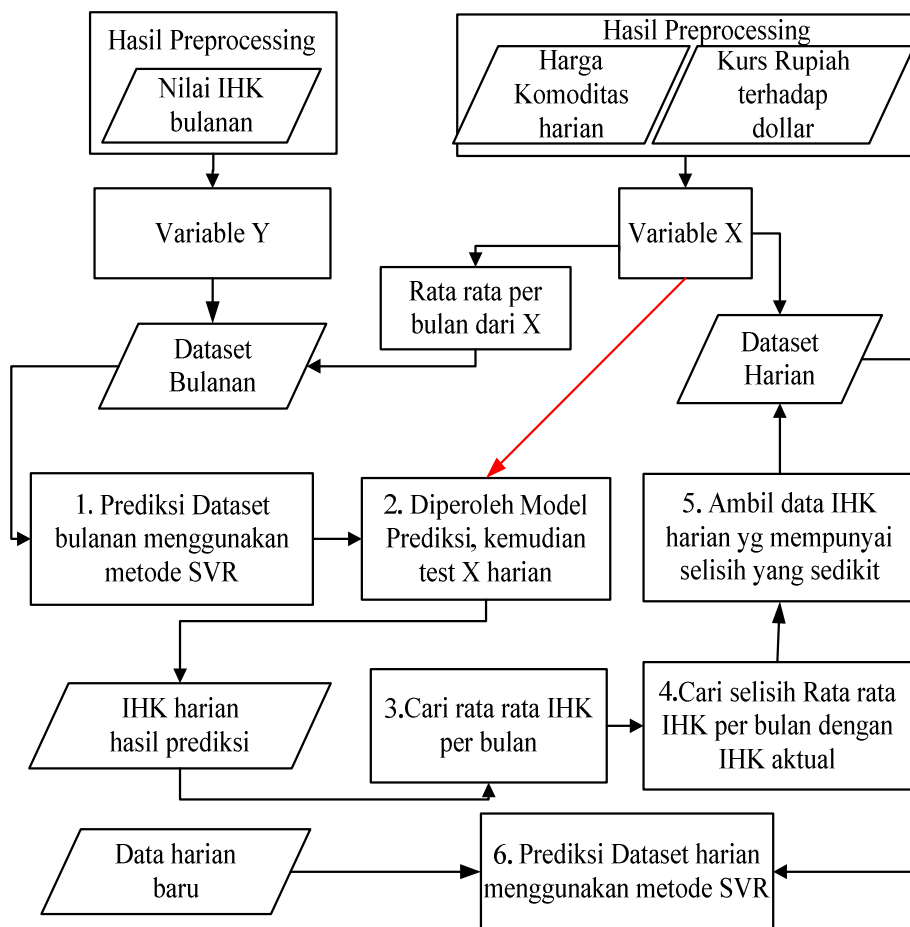
Tabel 3.8 Penerapan Grid Search pada Kernel Ridge Regression

		Gamma				
		0.001	0.1	1	10	100
Alpha	1	(0.001,1)	(0.1,1)	(1,1)	(10,1)	(100,1)
	0.1	(0.001,0.1)	(0.1,0.1)	(1,0.1)	(10,0.1)	(100,0.1)
	0.01	(0.001, 0.01)	(0.1,0.01)	(1,0.01)	(10,0.01)	(100,0.01)
	0.001	(0.001,0.001)	(0.1,0.001)	(1,0.01)	(10,0.001)	(100,0.001)

Setiap kombinasi Gamma dan Alpha akan di evaluasi nilai koefisien determinannya.

3.6.2 Pembangunan prediksi IHK bulanan

Pertama, dibangun prediksi data bulanan yang mempunyai akurasi baik. Dalam prediksi bulanan, penelitian ini menggunakan metode *Linear Regression*, *Kernel Ridge Regression* (KRR) dan *Support Vector Regression* (SVR). Penggunaan metode KRR bertujuan agar model prediksi lebih sensitif terhadap masalah *multicolinearity*[17] daripada metode *Linear Regression*. *Multicolinearity* berarti terjadi hubungan kuat antara variabel independent atau variabel masukan X. Jika terjadi hubungan kuat, model prediksi mempunyai akurasi yang kurang bagus. Ridge Regression mempunyai nilai alpha yang berfungsi sebagai penalti yang akan dikenakan pada data. Pada metode SVR, digunakan tiga kernel yaitu linear, RBF dan Poly.



Gambar 3.3 Diagram pembangunan prediksi IHK

Prediksi data bulanan ini digunakan untuk mengestimasi nilai IHK harian. Proses yang didalam pembangunan prediksi IHK bulanan ini ditunjukkan di dalam Gambar 3.3 diagram pembangunan prediksi IHK. Di dalam Gambar 3.3 dapat dilihat, setelah data nilai IHK bulanan dipreprocessing (disamakan tahun dasarnya) maka data ini menjadi variabel y. Data harga komoditas harian dan kurs rupiah terhadap dollar setelah dipreprocessing dijadikan variabel X harian. Kemudian dibentuk dataset harian dan dataset bulanan.

Dataset bulanan :

Variabel X = {rata-rata bulanan variabel X harian}

Variabel y = data nilai IHK bulanan

Untuk lebih menjelaskan Gambar 3.3, langkah tersebut dituliskan di dalam algoritma estimasi nilai IHK harian.

Algoritma estimasi IHK harian

Setelah dataset bulanan terbentuk ,

1. Melakukan pembangunan prediksi IHK dengan menggunakan data bulanan menggunakan Kernel Ridge Regression, Linear Regression, SVR (langkah no. 1 didalam Gambar 3.3).
2. Setelah model prediksi bulanan dihasilkan, pengujian variabel X harian dilakukan sehingga dihasilkan IHK harian.
3. Mencari rata-rata IHK per bulan
4. Mencari selisih rata rata IHK bulanan dengan nilai aktual IHK.
5. Pilih data IHK yang mempunyai selisih sedikit.

Setelah Algoritma menentukan IHK harian dijalankan, diperoleh dataset harian.

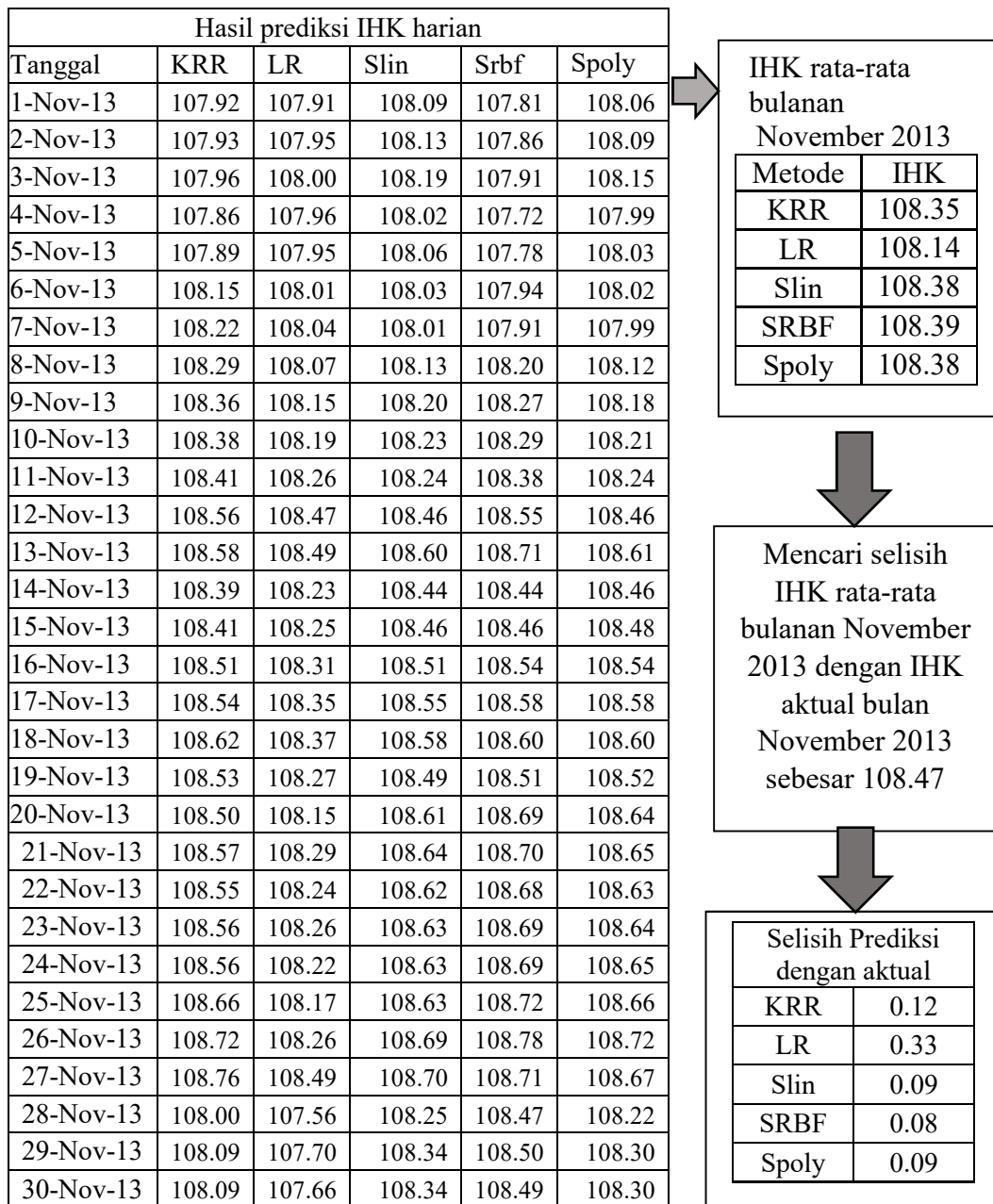
Dataset harian :

Variabel X = variabel X harian

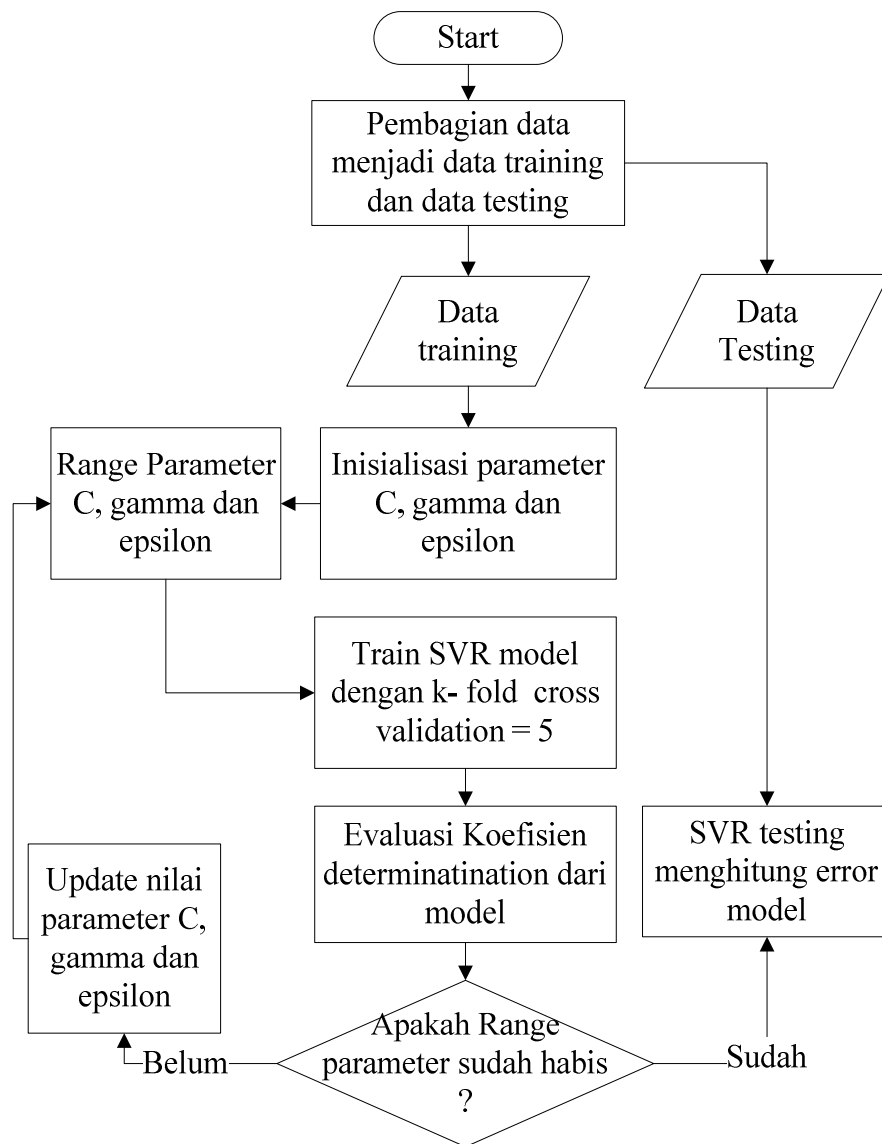
Variabel y = IHK hasil estimasi

Estimasi nilai IHK hari ini dijelaskan kembali dengan mengambil contoh data bulanan Kota Surabaya di bulan November 2013, ditunjukkan di dalam Gambar 3.4, proses estimasi IHK harian. Gambar 3.4 menunjukkan tabel data hasil prediksi variabel X harian dengan menggunakan lima metode prediksi. Kemudian data tersebut dicari rata bulanan untuk bulanan November 2013. Setelah itu dicari selisih antara data hasil prediksi IHK harian ini dengan nilai aktual IHK bulan November 2013.

Pembangunan model prediksi IHK ini juga menggunakan metode optimisasi hyperparameter. Di dalam Gambar 3.5 menunjukkan pemilihan hyperparameter menggunakan Grid search dan metode prediksi SVR. Dimulai dengan pembagian data menjadi data training dan data testing, inisiasi parameter C, gamma dan epsilon untuk menentukan parameter pertama yang akan digunakan dalam proses prediksi. Kemudian menentukan range parameter, range nilai parameter yang akan dikombinasikan di dalam model prediksi.



Gambar 3.4 Proses estimasi IHK harian



Gambar 3.5. Flowchart pembangunan model menggunakan optimisasi Grid Search

Langkah selanjutnya yaitu melakukan fitting data atau training data dengan cross validation sama dengan lima. Evaluasi model dilakukan dengan menggunakan metode koefisien determinat. Simpan nilai evaluasi koefisien determinat untuk dibandingkan dengan nilai evaluasi kombinasi lainnya. Nilai koefisien determinat dibandingkan dan dipilih nilai koefisien determinat yang mendekati nilai satu. Selama range parameter belum dikombinasikan semuanya, proses diatas tetap berulang. Hasil terakhir dari metode ini adalah kombinasi

hyperparameter dengan koefisien determinan hasil evaluasinya mendekati nilai satu.

Evaluasi Model

MSE dan R^2 merupakan metode evaluasi yang digunakan di dalam pembangunan prediksi ini. Dimana MSE di dalam persamaan 2.5 dan R^2 di dalam persamaan 2.6, untuk dataset ini maka

y = nilai IHK aktual dan

\hat{y}_i = nilai IHK hasil prediksi

Untuk $n_{samples}$ dataset yaitu 42, 48, 54, 1278, 1461, 164 volume dataset.

3.6.3 Pembangunan prediksi harian

Pembangunan model kedua yaitu pembangunan model prediksi harian. Pembangunan ini bertujuan membuat model prediksi IHK harian, mencari nilai estimasi IHK harian jika data harian yang baru datang. Prediksi ini menggunakan empat metode. Empat metode yang akan dibandingkan yaitu *Linear Regression*, *Kernel Ridge Regression*, *Support Vector Regression* dan *Random Forest Regression*. Di dalam tahap ini dilakukan proses *Optimisasi hyperparameter* untuk mencari hyperparameter terbaik. Kemudian ditunjukkan perbandingan waktu komputasi dari keempat metode prediksi dan juga kedua metode optimisasi hyperparameter.

3.7 Tool penelitian

Setelah dataset disiapkan, tahap berikutnya adalah mempersiapkan tool komputasi *python* di mesin dengan sistem Operasi *Ubuntu* 14.04.

Bahasa pemrograman : *python* v3.6.1, alasan pemrograman ini dipilih karena kemudahannya, multifungsi (banyak modul dukungannya) dan bisa digunakan untuk analitik data. Berjalan di banyak sistem operasi.

Interface yang digunakan : antar muka web menggunakan *jupyter notebook*. *Jupyter notebook* berjalan diatas pemrograman *python*.

Library (modul) dari *python* yang digunakan dalam penelitian ini adalah :

- Numpy

- SciPy
- Scikit-learn
- Matplotlib

3.8 Skenario Simulasi Cloud Computing

Di dalam migrasi prediksi IHK harian ke dalam cloud computing, penelitian ini merujuk ke framework Cloud-RMM yang telah sebutkan di Table 2.4, framework migrasi cloud-RMM. Penggunaan framework dibatasi. Di dalam penelitian ini hanya membahas :

1. Proses perencanaan

Studi kelayakan dari cloud, analisa kebutuhan migrasi, pemilihan penyedia cloud dan layanan cloud, strategi migrasi.

2. Proses Pelaksanaan

Modifikasi kode pemrograman aplikasi, ekstrasi arsitektur, ekstrasi data, transformasi

3. Evaluasi

Pengembangan, Pengujian dan validasi.

Proses perencanaan dibahas dalam sub bab ini dan proses pelaksanaan akan dibahas di Bab Hasil dan Pembahasan. Di tahap perencanaan ini akan dibahas apa saja yang disiapkan sebelum proses migrasi dilakukan meliputi :

- a. Studi kelayakan;

Di dalam bab Kajian teori subbab 2.4 Cloud Computing, telah dibahas sekilas tentang cloud computing dan penyedia cloud.

- b. Analisa kebutuhan migrasi

Model prediksi ini yaitu model prediksi IHK membutuhkan :

- Mesin komputasi dengan sistem operasi Ubuntu versi minimal 14.04, pemrograman python dan interface jupyter notebook.
- Personal Komputer yang digunakan core i3 dengan memory 12 GB
- Penyedia cloud computing yang digunakan dua yaitu Amazon Web Service (AWS) dan Microsoft Azure.

- Model layanan cloud yang digunakan dalam penelitian ini adalah IAAS (Infrastructure as Service). Dalam IAAS, pengguna seperti mempunyai mesin komputasi sendiri (lihat bab 2.4). Mesin komputasi ini diwujudkan dalam bentuk *Virtual Machine* yang bisa diremote atau diakses dari luar (pengguna dan mesin tidak berada di dalam satu area). AWS menyediakan layanan IAAS dengan nama Amazon Elastic Compute Cloud atau dikenal dengan nama (EC2). Penelitian ini menggunakan akun tidak berbayar, sehingga ada keterbatasan penelitian di dalam cloud computing. Keterbatasan didalam memanfaatkan fungsi cloud computing *scalability*.

Perbedaan pelayanan yang diberikan oleh akun gratis cloud AWS dan Azure ditunjukkan di dalam Tabel 3.8 yaitu tabel Perbedaan akun gratis AWS dan Azure.

Tabel 3.9 Perbedaan akun gratis AWS dan Azure

	AWS	Azure
Masa free account	12 bulan	1 bulan dengan free credit Rp.2.000.000,-
Spesifikasi VM	1 CPU, memory 1 GB, Harddisk 8GB	1 CPU memory 3.5GB
	Tidak diperbolehkan mengakses AWS marketplace yang berisi aplikasi instant. Dengan aplikasi instant tersebut, pengguna tidak perlu menginstall program di virtual machinenya	Diperbolehkan menginstall paket instalasi yang berisi aplikasi data analytic sehingga tidak perlu lagi melakukan instalasi python dan jupyter notebook

c. Strategi migrasi

Simulasi prediksi IHK harian di lingkungan Cloud Computing menggunakan pemrograman python dengan interface *jupyter notebook*. Langkah yang dilakukan adalah :

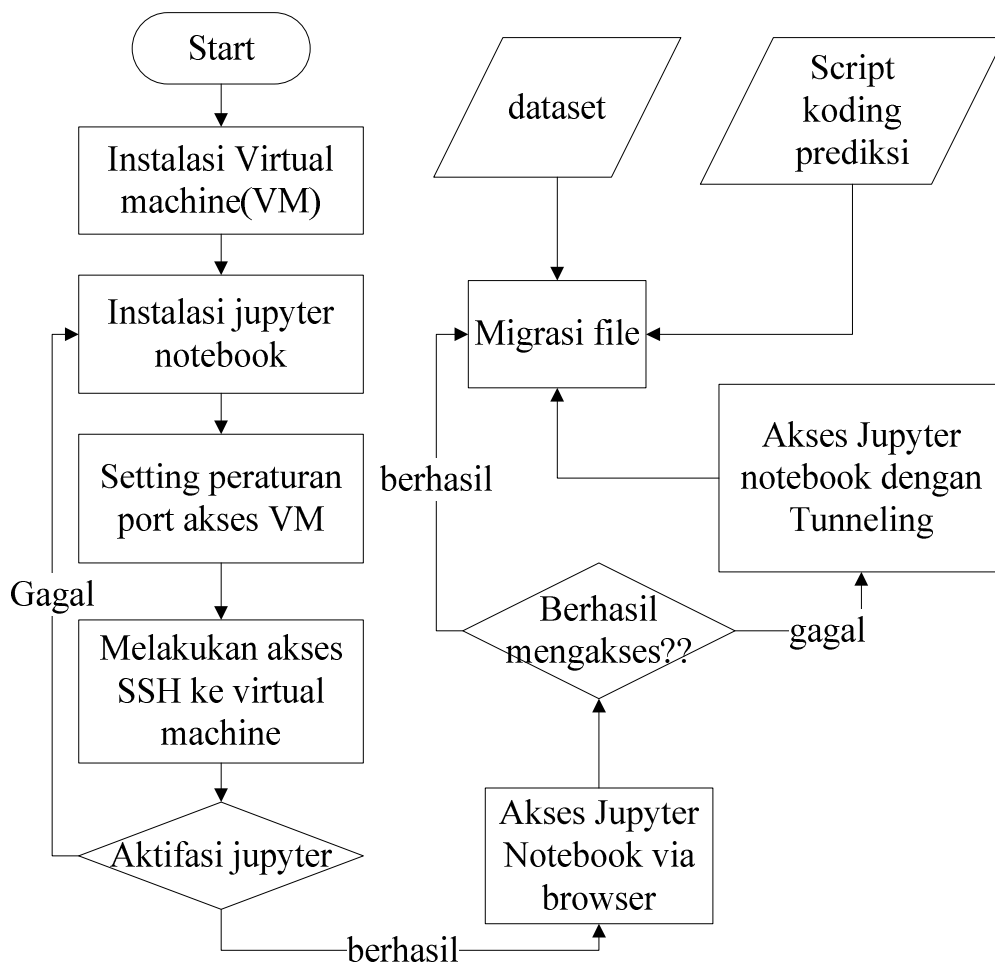
1. Melakukan instalasi *Virtual Machine* (VM) dengan sistem operasi *Ubuntu* versi 14.04. Kelebihan dari pembangunan aplikasi di cloud computing, penyedia cloud sudah memberikan pilihan sistem operasi bahkan tersedia juga pilihan paket misalnya paket analitik. User bisa menginstal *VM* dan memilih sistem operasi *virtual machine* tersebut sesuai kebutuhan.
2. Instalasi *python* versi 3.6 dan *jupyter notebook*.
3. Sebelum pengguna mengakses *jupyter notebook* melalui browser, perlu dilakukan pengaturan *inbound* dan *outbound* rule. *Inbound* dan *outbound* rule ini merupakan aturan untuk mengatur keamanan dari *virtual machine*. Aturan tersebut membatasi akses dari *virtual machine*, apakah *virtual machine* bisa diakses melewati port tersebut atau tidak, dan pelayanan apa saja yang bisa diakses melewati port. Rule ini harus diterapkan agar *jupyter notebook* dapat diakses dari luar. Dalam *jupyter notebook*, port yang biasa digunakan adalah 8888 dengan service port TCP. Nomor *port* ini bisa diubah melalui file konfigurasi dari python.
4. Setelah pengaturan selesai, aktifasi *jupyter notebook* di *virtual machine*. *Jupyter notebook* dapat diakses melalui penjelajah web atau *web browser* dengan format *https://ip-publik:port*.

Terkadang ada kendala di dalam mengakses *jupyter notebook* secara remote. Ini dikarenakan aturan atau *firewall* dari *virtual machine*. Jika sudah dapat dipastikan bahwa *inbound/ outbound* rule disetting dengan benar, maka perlu dilakukan cara lain, yaitu membuka jalur akses *notebook* dengan cara *tunneling*. Gambar 3.6 menunjukkan tahapan migrasi prediksi IHK ke lingkungan cloud computing seperti yang telah dijelaskan dalam langkah di atas.

Skenario pengujian simulasi prediksi IHK harian di dalam cloud yaitu :

1. Evaluasi prosedur dan fungsionalitas model di Personal Computer, Amazon EC2 dan di Azure
2. Evaluasi nilai akurasi model di Personal Computer, Amazon EC2 dan di Azure

3. Evaluasi waktu training dan waktu testing prediksi IHK di Personal Computer, Amazon EC2 dan di Azure
4. Evaluasi nilai akurasi prediksi IHK di Personal Computer, Amazon EC2 dan di Azure



Gambar 3.6 Tahapan migrasi prediksi IHK ke cloud computing

3.8.1 Rancangan evaluasi prosedur dan fungsionalitas.

Evaluasi pengujian fungsionalitas model dilakukan di Personal Computer, Amazon EC2 dan di Azure. Tujuan dari evaluasi ini adalah menguji fungsionalitas setiap model prediksi jika dijalankan di dalam Personal Komputer

dan di Cloud Computing. Seperti halnya sistem, jika dimigrasikan ke dalam lingkungan komputasi lainnya harus bisa berjalan, jika terjadi kendala harus sudah diperhitungkan sebelum terjadi migrasi. Prosedure migrasi yang telah ditunjukkan didalam Gambar 3.6, Tahapan migrasi prediksi IHK ke cloud computing, diterapkan pada personal computer (PC), Amazon EC2 dan di Azure. Terkecuali pada PC, tidak dilakukan Instalasi Virtual Machine karena model prediksi IHK tersebut dibangun di localhost PC. Tabel 3.9 menunjukkan Rancangan evaluasi prosedur dan fungsionalitas, setiap fungsi di dalam Gambar 3.6 dijalankan dan akan dilakukan pengujian fungsi manakah yang Gagal dijalankan, atau berhasil dilakukan di ketiga lingkungan komputasi.

Tabel 3.10 Rancangan evaluasi prosedur dan fungsionalitas

No	Fungsi	PC	EC2	Azure
1	Instalasi Virtual Machine			
2	Instalasi Python			
3	Instalasi <i>Jupyter notebook</i>			
4	Akses SSH ke virtual Machine via terminal			
5	Akses <i>Jupyter Notebook</i> via browser			
6	Migrasi File Script dan dataset			
7	Akses dataset dari <i>jupyter notebook</i>			

Nilai dari pengujian diatas adalah :

- **Berhasil** : jika pengujian berhasil dilaksanakan tanpa ada notifikasi gagal.
- **Gagal** : jika pengujian tidak berhasil dilaksanakan

3.8.2 Rancangan evaluasi waktu komputasi metode optimisasi

Evaluasi metode Grid Search dan Random Search yang dilakukan di dalam penelitian ini yaitu dengan cara membandingkan waktu proses yang dibutuhkan dalam penggunaan metode Grid Search dan Random Search. Selain

itu, dalam evaluasi ini ditunjukkan nilai koefisien determinan R^2 yang dihasilkan kedua metode tersebut. Tujuan dari evaluasi ini untuk mengevaluasi apakah komputasi yang berjalan di cloud lebih baik daripada di personal komputer. Perbandingan dilakukan jika dijalankan di dalam PC, Amazon EC2 dan Azure. Rancangan evaluasi waktu komputasi untuk metode optimisasi hyperparameter di cloud computing, ditunjukkan dalam Tabel 3.10, tabel Rancangan evaluasi penggunaan metode Grid dan Random Search.

Data yang digunakan dalam pengujian adalah data bulanan dan data harian kota Surabaya. Pengujian dilakukan dengan volume data atau n yang beragam yaitu 42, 48, 54, 1278 1461 dan 1644 data. Variasi volume ini untuk menunjukkan bagaimana perbandingan waktu komputasi dalam memproses data yang volumenya makin lama makin besar.

Tabel 3.11 Rancangan evaluasi penggunaan metode Grid dan Random Search

No	n	Metode	Personal Computer				Amazon EC2				Azure			
			Grid		Random		Grid		Random		Grid		Random	
			t	R^2	t	R^2	t	R^2	t	R^2	t	R^2	t	R^2
1	n	SVR-Linear												
		SVR-RBF												
		SVR-Poly												
		Random Forest												

Keterangan :

t = waktu komputasi yang dibutuhkan dalam satuan detik

R^2 = nilai koefisien determinan yang diperoleh

n = jumlah data sebesar 42, 48, 54, 1278, 1461, 1644

3.8.3 Rancangan evaluasi waktu proses training dan testing

Dalam evaluasi ini, dilakukan proses pelatihan (training) dan pengujian (testing) menggunakan dua metode prediksi yaitu SVR dan Random Forest. Data yang diujikan tetap menggunakan data dari Kota Surabaya. Proses pelatihan dan pengujian tersebut menggunakan hyperparameter terbaik hasil proses Grid Search. Kemudian dilakukan pengukuran waktu training dan testing yang dibutuhkan jika dijalankan di PC, Amazon EC2 dan Azure. Evaluasi ini bertujuan untuk menunjukkan perbandingan waktu komputasi dua metode prediksi SVR dan Random Forest, dan perbandingan komputasi jika berjalan di dalam Cloud Computing. Dalam Tabel 3.11 menunjukkan Rancangan evaluasi waktu proses training dan testing data, menunjukkan waktu training dan testing di Personal Komputer, Azure dan Amazon EC2 untuk volume 42, 48, 54, 1278, 1461, 1644 dataset.

Tabel 3.12 Rancangan evaluasi waktu proses training dan testing data

No	n	Metode	Personal Computer		Azure		Amazon EC2	
			Training	Testing	Training	Testing	Training	Testing
1	n	SVR-Linear						
		SVR-RBF						
		SVR-Poly						
		Random Forest						

Keterangan :

n = jumlah data sebesar 42, 48, 54, 1278, 1461, 1644

Kolom Training berisi waktu training dalam satuan detik

Kolom Testing berisi waktu testing dalam satuan detik

3.8.4 Rancangan evaluasi nilai akurasi prediksi IHK

Evaluasi model diperlukan untuk memastikan model yang telah dibangun merupakan model yang mempunyai performansi akurasi yang baik (ditandai dengan MSE dan R^2) dan waktu komputasi yang cukup singkat. Evaluasi model regresi ini menggunakan metode evaluasi MSE yang ada di dalam persamaan (2.5) dan metode evaluasi R^2 (coefficient of determination) yang ada di dalam persamaan (2.6). Nilai keduanya diperoleh dengan membandingkan nilai target aktual y dengan nilai hasil prediksi \hat{y} .

Model mempunyai akurasi yang bagus, jika nilai MSE lebih kecil atau mendekati nilai nol. Sedangkan nilai R^2 yang baik jika nilai R^2 mendekati nilai satu. Nilai terbaik R^2 ini adalah satu. Nilai R^2 negative bisa terjadi, menandakan model masih belum optimal [19]

Evaluasi model prediksi IHK menggunakan nilai Mean Square Error/ MSE dan koefisien Determinant atau R^2 . Di dalam pengujian ini akan ditunjukkan konsistensi model prediksi jika dijalankan di tiga sumber komputasi yaitu personal komputer, Azure dan EC2, bisa ditunjukkan dengan Tabel 3.12 Rancangan evaluasi nilai MSE dan R^2 . Model prediksi IHK yang telah diimplementasikan di ketiga sumber akan diujikan, dan dievaluasi apakah model konsisten menghasilkan nilai evaluasi model, MSE dan R^2 yang sama atau tidak.

Tabel 3.13 Rancangan evaluasi nilai MSE dan R^2

No	n	Metode	Personal Computer		Azure		Amazon EC2	
			MSE	R^2	MSE	R^2	MSE	R^2
1	n	SVR-Linear						
		SVR-RBF						
		SVR-Poly						
		RandomForest						

Keterangan :

N = jumlah data sebesar 42, 48, 54, 1278, 1461, 1644

BAB 4

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan tool eksperimen *framework python* dengan antarmuka berbasis web menggunakan *shell python* atau dikenal dengan *jupyter notebook* agar mudah diimplementasikan di Big Data Technology seperti *SPARK* karena *SPARK* menggunakan bahasa *python* dan *scala* dan juga bisa dengan mudah dimigrasikan ke lingkungan cloud computing.

4.1 Analisa awal data

Di dalam bab 3 Metodologi Penelitian, telah dibahas bagaimana variable yang digunakan dalam prediksi ini dipilih. Dalam bab hasil dan pembahasan ini, dibahas bagaimana mengatasi hilangnya nilai dalam data setelah variabel prediksi dipilih. Data online seringkali mengalami kesalahan nilai masukan disebabkan oleh *human error*. Oleh karena itu diperlukan *preprocessing* untuk memproses data sebelum digunakan dalam model prediksi.

Dalam Tabel 4.1 ditunjukkan data harga komoditas kota Surabaya variabel Daging Sapi murni, Daging Ayam Broiler dan Daging Ayam Kampung pada tanggal 20 November s.d 30 November 2012.

Tabel 4.1 Nilai hilang dalam dataset harian kota Surabaya

Tanggal	Daging Sapi Murni	Daging Ayam Broiler	Daging Ayam Kampung
20-Nov-12	75200	22100	42600
21-Nov-12	77100	22100	42600
22-Nov-12	78500	21500	41800
23-Nov-12	79800	21420	41800
24-Nov-12	0	20200	43000
25-Nov-12	0	20360	43200
26-Nov-12	0	20960	43800
27-Nov-12	0	21000	43800
28-Nov-12	0	20800	42600
29-Nov-12	82500	20060	41800
30-Nov-12	86500	20060	41800

Terdapat nilai yang hilang pada variabel Daging Sapi Murni pada tanggal 24 November 2012 s.d 28 November 2012 (Tabel 4.1 ditandai blok cell yang diberi kotak). Di tanggal sebelum 24 Nov dan setelah 28 Nov, variabel daging terdapat nilai variabelnya. Pada tanggal 24.s.d 28 Nov variabel Daging Ayam Broiler dan Daging Ayam Kampung terisi nilainya.

Tabel 4.2 Dataset harian bulan November 2012 di Provinsi Jawa Timur

Tanggal	Daging Sapi Murni	Daging Ayam Broiler	Daging Ayam Kampung
20-Nov-12	75117	22127	41600
21-Nov-12	75562	21925	41427
22-Nov-12	76014	21588	41561
23-Nov-12	76126	21453	41379
24-Nov-12	73446	21482	41670
25-Nov-12	73529	21673	41663
26-Nov-12	75399	21949	41523
27-Nov-12	75920	21799	41488
28-Nov-12	76350	21918	41377
29-Nov-12	79706	21792	40998
30-Nov-12	80104	21792	41028

Didalam data tersebut tidak memungkinkan menggunakan nilai rata-rata daerah, yaitu nilai daging sapi tanggal 24 – 28 November 2012 di Jawa Timur. Di dalam Tabel 4.2 menunjukkan dataset harian untuk variabel Daging Sapi murni, Daging Ayam Broiler dan Daging Ayam Kampung pada tanggal 20 November s.d 30 November 2012. Data Jawa Timur ini diperoleh dari rata – rata nilai variabel di 38 kota / kabupaten di Jawa Timur. Nilai rentang Jawa Timur pada tanggal 23 November 2012 sebesar Rp.73.446,- dan tgl 29 November sebesar Rp.79.706,-. Sedangkan pada tanggal 23 November, harga daging di kota Surabaya Rp.79.800,- dan tanggal 29 November Rp. 82.500,-. Jadi, harga daging sapi rata-rata di Jawa Timur lebih rendah daripada di Surabaya. Oleh karena itu, untuk mengisi kekosongan nilai variabel daging di kota Surabaya pada tanggal 24 November 2012 s.d 28 November 2012 menggunakan nilai yang sama dengan hari sebelum kekosongan terjadi.

Di dalam Tabel 4.3 ditunjukkan Hasil analisa Telur Ayam Kampung dari kota Malang untuk data tahun 2016. Dari data Tabel 4.3 ditemukan perbedaan

range harga yang cukup signifikan di bulan Februari tahun 2016. Di dalam kolom data komoditas telur ayam kampung, pada tanggal 4 Februari 2016, yang ditandai dengan warna hijau, harga telur ayam kampung sebesar Rp. 36.000,- , tanggal 26 Februari 2016 sebesar Rp. 65.000,- sedangkan pada tanggal 19, 20, 25 dan 26 Mei 2016 harga telur ayam kampung sebesar Rp. 2000,- Perbedaan harga tanggal 4 Februari dengan 26 Mei terlalu jauh dan tidak memungkinkan terjadi. Penurunan nilai Rp. 65.000 menjadi Rp. 2000,-. Sebagai perbandingan, pada tanggal 4 Februari 2016 harga rata rata Telur ayam kampung di Jawa Timur sebesar Rp. 32.139,- tanggal 21 Februari 2016 Rp. 35.553,- sedangkan pada tanggal 26 Mei 2016 harga rata rata di Jawa Timur Rp. 33.553,- Selisih yang besar ini bisa mempengaruhi hasil prediksi IHK. Nilai telur ayam kampung di Kota Malang pada tanggal 21 Februari dan 19, 20, 25 dan 26 Mei 2016 perlu ditinjau ulang, apakah terjadi kesalahan dari operator aplikasi daerah atau pada kenyataannya, nilai itu terjadi.

Tabel 4.3 Analisa harga telur ayam kampung di dataset Kota Malang

Tanggal	Telur Ayam Ras / Petelur	Telur Ayam Kampung
1-Feb-16	22600	36000
2-Feb-16	22600	36000
3-Feb-16	22600	36000
4-Feb-16	22600	36000
21-Feb-16	20900	65000
18-May-16	18700	0
19-May-16	18800	2000
20-May-16	18800	2000
21-May-16	18800	0
22-May-16	18800	0
23-May-16	19000	0
24-May-16	18800	0
25-May-16	18800	2000
26-May-16	18940	2000
27-May-16	18940	0

Hasil analisa data Kota Kediri, terdapat banyak kehilangan nilai data. Warna kuning di Tabel 4.4 yaitu data kota Kediri Tahun 2016 yang mengalami kekosongan nilai, menunjukkan bahwa data pada tanggal 15, 16, 25, 28

November 2016 dan 22, 23, 28, 31 Desember 2016 mengalami kekosongan nilai. Untuk mengatasi permasalahan ini ada beberapa cara yang bisa dilakukan yaitu menghapus data pada tanggal tersebut, mengisi nilai pada tanggal tersebut dengan nilai sebelumnya, atau mengisi nilai pada tanggal tersebut dengan nilai mediannya. Selain masalah diatas, terdapat juga data yang salah memasukan, Rp.10.000,- (diasumsi Rp. 10.000 karena melihat nilai data pada tanggal sebelumnya) di masukkan menjadi 10,-

Tabel 4.4 Data kota Kediri tahun 2016 yang mengalami kekosongan nilai

Tanggal	- Bengawan	- Mentik	- IR 64	- Gula Pasir Dalam Negri
14-Nov-16	9917	10000	8833	12000
15-Nov-16	0	0	0	0
16-Nov-16	0	0	0	0
17-Nov-16	10000	10000	8833	12000
24-Nov-16	9750	10000	8967	12167
25-Nov-16	0	0	0	0
26-Nov-16	9750	10000	8967	12167
27-Nov-16	9625	10000	8950	12250
28-Nov-16	0	0	0	0
29-Nov-16	9750	10000	8967	12417
21-Dec-16	9833	10000	9050	12417
22-Dec-16	0	0	0	0
23-Dec-16	0	0	0	0
24-Dec-16	9833	10000	9050	12417
27-Dec-16	9833	10000	9050	12500
28-Dec-16	0	0	0	0
29-Dec-16	9833	10000	9050	12333
30-Dec-16	10000	9833	9050	12333
31-Dec-16	0	0	0	0

Normalisasi data ini diperlukan untuk memperoleh dataset yang baik sehingga nilai akurasi model prediksi yang didapatkan juga tinggi. Karena banyak terjadi nilai yang hilang di 8 kota perwakilan survey BPS, didalam penelitian ini, model prediksi IHK hanya menggunakan data Kota Surabaya karena mempunyai nilai hilang yang kecil, yaitu 5. Nilai komoditas yang hilang di 8 kota perwakilan survey BPS dirangkum di dalam Tabel 4.5 yaitu tabel hasil ringkasan data yang hilang.

Tabel 4.5 Hasil ringkasan nilai yang hilang pada 8 kota di Jawa Timur

No	Kota	Nilai yang hilang
1	Kota Surabaya	5
2	Kota Malang	1578
3	Kota Kediri	319
4	Kota Probolinggo	1043
5	Kota Madiun	1329
6	Kab. Sumenep	627
7	Kab. Jember	467
8	Kab Banyuwangi	1596

Analisa variable nilai tukar rupiah terhadap dollar memperoleh hasil bahwa variable nilai tukar rupiah terhadap dollar diupdate nilainya tiap hari kerja. Oleh karena itu, untuk mengisi kekosongan data pada hari libur, digunakan nilai Median dari nilai tukar rupiah terhadap dollar.

4.2 Hasil Preprocessing

Data IHK bulanan di Surabaya ditunjukkan di dalam Tabel 4.6, tabel Perubahan IHK dengan basis tahun yang sama. Nilai IHK bulanan pada bulan januari 2012 sampai dengan Desember 2013 menggunakan tahun dasar 2007 sedangkan nilai IHK bulan Januari tahun 2014 sampai dengan bulan Desember tahun 2016 menggunakan tahun dasar 2012. IHK bulanan pada tahun 2012 s.d 2014 diubah nilainya mengikuti tahun dasar 2012 dengan menggunakan rumus (3.1) sedangkan data IHK tahun 2014 – 2016 tidak mengalami perubahan.

4.3 Hasil pembangunan prediksi menggunakan data bulanan

Pembangunan prediksi IHK dilakukan menggunakan pemrograman python. Prediksi bulanan dibangun dengan 60 record dari data kota Surabaya dan terbagi menjadi 42 data training dan 18 data testing. Dari prediksi bulanan ini akan dilakukan estimasi nilai IHK harian. Algoritma prediksi IHK dan inisiasi parameternya telah dijelaskan di bab tiga. Dalam pembangunan prediksi bulanan ini menggunakan metode Linear regression, Kernel Ridge Regression dan Support Vector Regression(SVR). SVR menggunakan tiga kernel yaitu Linear, RBF dan

Polynomial. Hasil dari optimisasi *hyperparameter* Grid Search ditampilkan dalam Tabel 4.7, Hasil optimisasi *hyperparameter* menggunakan Metode Grid Search.

Tabel 4.6. Perubahan IHK dengan basis tahun yang sama di Kota Surabaya

Bulan	Nilai IHK	Nilai inflasi	IHK perubahan	Bulan	Nilai IHK	Nilai inflasi	IHK perubahan
Jan-12	133.63	0.1	98.2	Jul-14	112.23	0.42	112.23
Feb-12	134.11	0.36	98.55	Aug-14	112.79	0.5	112.79
Mar-12	134.4	0.22	98.77	Sep-14	113.25	0.41	113.25
Apr-12	134.72	0.24	99.01	Oct-14	113.8	0.49	113.8
May-12	134.51	-0.16	98.85	Nov-14	115.24	1.27	115.24
Jun-12	135.18	0.5	99.34	Dec-14	117.81	2.23	117.81
Jul-12	136.16	0.72	100.06	Jan-15	118.29	0.41	118.29
Aug-12	137.7	1.13	101.19	Feb-15	117.79	-0.42	117.79
Sep-12	137.48	-0.01	101.03	Mar-15	118.69	0.41	118.69
Oct-12	137.71	0.17	101.2	Apr-15	119.15	0.39	119.15
Nov-12	138.18	0.34	101.54	May-15	119.79	0.54	119.79
Dec-12	140.1	1.39	102.95	Jun-15	120.25	0.38	120.25
Jan-13	141.15	0.75	103.73	Jul-15	120.83	0.48	120.83
Feb-13	142.52	0.97	104.73	Aug-15	121.14	0.26	121.14
Mar-13	141.99	-0.37	104.34	Sep-15	120.73	-0.34	120.73
Apr-13	140.98	-0.71	103.6	Oct-15	120.71	-0.02	120.71
May-13	142.08	0.78	104.41	Nov-15	121.85	0.94	121.85
Jun-13	147.2	3.6	108.17	Dec-15	122.74	0.73	122.74
Jul-13	148.55	0.92	109.17	Jan-16	122.6	-0.11	122.6
Aug-13	147.43	-0.75	108.35	Feb-16	122.67	0.06	122.67
Sep-13	147.73	0.2	108.56	Mar-16	122.49	-0.15	122.49
Oct-13	147.59	-0.09	108.47	Apr-16	122.65	0.13	122.65
Nov-13	148.57	0.66	109.18	May-16	123.5	0.69	123.5
Dec-13	110.47	1.18	110.47	Jun-16	124.53	0.83	124.53
Jan-14	110.72	0.23	110.72	Jul-16	124.65	0.1	124.65
Feb-14	110.97	0.23	110.97	Aug-16	124.88	0.18	124.88
Mar-14	111.16	0.17	111.16	Sep-16	124.75	-0.1	124.75
Apr-14	111.35	0.17	111.35	Oct-16	125.07	0.26	125.07
May-14	111.76	0.37	111.76	Nov-16	125.77	0.56	125.77

Metode Optimisasi menghasilkan hyperparameter yang terbaik yang dapat digunakan untuk proses pelatihan dan pengujian di prediksi IHK. Di bab 3.61 telah disajikan range parameter yang akan dicari kombinasi terbaik. Range parameter yang digunakan di dalam penelitian ini, disajikan di dalam Tabel 4.7.

Tabel 4.7 Range hyperparameter yang digunakan

Metode	Alpha	Gamma	epsilon	C	degree
Kernel Ridge	1, 0.1, 0.01, 0.001	0.01, 0.1, 1, 10			
Linear Regression					
SVR-linear		0.1, 1.1, 2.1	0.1, 1.1, 2.1	1 s.d 99	
SVR-RBF		0.1, 1.1, 2.1	0.1, 1.1, 2.1	1 s.d 99	
SVR-Poly		0.1, 1.1, 2.1	0.1, 1.1, 2.1	1 s.d 99	1,2,3

Dari range parameter di dalam Tabel 4.7, kolom alpha, Gamma, epsilon, C dan degree, dicari kombinasi terbaik menggunakan metode Grid Search. Grid search menghasilkan score yang mewakili nilai koefisien determinan R^2 . Nilai score yang tinggi berarti kombinasi yang diperoleh bagus. Hasil koefisien determinan tersebut disajikan di dalam Tabel 4.8. Data di dalam Tabel 4.8 menyajikan nilai hyperparameter C dengan range 1 sampai dengan 54 dan dengan range gamma sebesar 0.1 dan 1.1 dan kolom epsilon yang bernilai 0.1. Di dalam fungsi Grid Search dari library Scikit python, selain menghasilkan nilai score (ditunjukkan di dalam Tabel 4.8 kolom R^2), dihitung juga ranking dari score tersebut (ditunjukkan di dalam Tabel 4.8 kolom Rank. Score yang telah dihasilkan diurutkan dari terbesar sampai terkecil, kemudian diberikan nilai rankingnya. Nilai R^2 yang besar memperoleh nilai ranking sebesar 1, nilai R^2 terbesar kedua diberikan nilai ranking 2 begitu seterusnya sampai semua nilai C (di dalam Tabel 4.8 ditunjukkan oleh kolom C) diberikan nilai ranknya. Begitu juga dengan parameter gamma dan epsilon. Di dalam Tabel 4.8, hyperparameter C yang memperoleh nilai ranking 1 adalah hyperparameter C dengan nilai sebesar 47, gamma dengan nilai sebesar 0.1 dan epsilon dengan nilai sebesar 0.1. Data lengkap hasil dari proses Grid Search disajikan lengkap sesuai range parameter yang telah dimasukkan di dalam Lampiran buku.

Tabel 4.8 Hasil R^2 dari Grid Search metode SVR-rbf

epsilon	gamma	C	rank	score	epsilon	gamma	C	rank	score
0.1	0.1	1	868	0.659248	0.1	1.1	1	874	0.409107
0.1	0.1	2	474	0.922429	0.1	1.1	2	767	0.771616
0.1	0.1	3	191	0.960257	0.1	1.1	3	481	0.897126
0.1	0.1	4	187	0.97247	0.1	1.1	4	471	0.926866
0.1	0.1	5	159	0.976199	0.1	1.1	5	380	0.938778
0.1	0.1	6	93	0.977402	0.1	1.1	6	378	0.943653
0.1	0.1	7	92	0.978593	0.1	1.1	7	375	0.946427
0.1	0.1	8	91	0.979744	0.1	1.1	8	372	0.948247
0.1	0.1	9	90	0.980436	0.1	1.1	9	370	0.949769
0.1	0.1	10	89	0.980755	0.1	1.1	10	365	0.951
0.1	0.1	11	86	0.980887	0.1	1.1	11	280	0.952025
0.1	0.1	12	88	0.980805	0.1	1.1	12	279	0.952917
0.1	0.1	13	87	0.98082	0.1	1.1	13	277	0.953397
0.1	0.1	14	85	0.981077	0.1	1.1	14	276	0.95356
0.1	0.1	15	84	0.981165	0.1	1.1	15	275	0.953707
0.1	0.1	16	83	0.981283	0.1	1.1	16	274	0.95383
0.1	0.1	17	82	0.98143	0.1	1.1	17	273	0.953943
0.1	0.1	18	81	0.981563	0.1	1.1	18	272	0.953958
0.1	0.1	19	80	0.981679	0.1	1.1	19	201	0.953975
0.1	0.1	20	79	0.981786	0.1	1.1	20	199	0.953988
0.1	0.1	21	78	0.981815	0.1	1.1	21	198	0.953997
0.1	0.1	22	77	0.981853	0.1	1.1	22	196	0.95401
0.1	0.1	23	76	0.981871	0.1	1.1	23	195	0.954015
0.1	0.1	24	75	0.981881	0.1	1.1	24	194	0.954022
0.1	0.1	25	74	0.981972	0.1	1.1	25	193	0.954028
0.1	0.1	26	73	0.982091	0.1	1.1	26	192	0.954031
0.1	0.1	27	66	0.982297	0.1	1.1	27	197	0.954005
0.1	0.1	28	58	0.982485	0.1	1.1	28	200	0.95398
0.1	0.1	29	53	0.982593	0.1	1.1	29	202	0.953962
0.1	0.1	30	49	0.982691	0.1	1.1	30	203	0.953962
0.1	0.1	31	46	0.982762	0.1	1.1	31	203	0.953962
0.1	0.1	32	43	0.98283	0.1	1.1	32	203	0.953962
0.1	0.1	33	41	0.982894	0.1	1.1	33	203	0.953962
0.1	0.1	34	38	0.982955	0.1	1.1	34	203	0.953962
0.1	0.1	35	36	0.983014	0.1	1.1	35	203	0.953962
0.1	0.1	36	31	0.983069	0.1	1.1	36	203	0.953962
0.1	0.1	37	28	0.983125	0.1	1.1	37	203	0.953962
0.1	0.1	38	26	0.983182	0.1	1.1	38	203	0.953962
0.1	0.1	39	24	0.983211	0.1	1.1	39	203	0.953962
0.1	0.1	40	22	0.983247	0.1	1.1	40	203	0.953962
0.1	0.1	41	18	0.983305	0.1	1.1	41	203	0.953962
0.1	0.1	42	15	0.983362	0.1	1.1	42	203	0.953962
0.1	0.1	43	12	0.98342	0.1	1.1	43	203	0.953962
0.1	0.1	44	8	0.98347	0.1	1.1	44	203	0.953962
0.1	0.1	45	6	0.983516	0.1	1.1	45	203	0.953962
0.1	0.1	46	4	0.983555	0.1	1.1	46	203	0.953962
0.1	0.1	47	1	0.983583	0.1	1.1	47	203	0.953962
0.1	0.1	48	2	0.983582	0.1	1.1	48	203	0.953962
0.1	0.1	49	3	0.983561	0.1	1.1	49	203	0.953962
0.1	0.1	50	5	0.983537	0.1	1.1	50	203	0.953962
0.1	0.1	51	7	0.983505	0.1	1.1	51	203	0.953962

Dapat disimpulkan parameter terbaik yang dihasilkan dari metode Grid search (di dalam Tabel 4.8 ditandai dengan kotak berwarna merah) untuk SVR kernel rbf adalah hyperparameter C sebesar 47. Sedangkan nilai hyperparameter gamma terbaik sebesar 0.1 dan epsilon terbaik sebesar 0.1

Data di dalam Tabel 4.9 menyajikan ringkasan hasil dari optimisasi hyperparameter menggunakan metode Grid Search untuk metode Kernel Ridge, SVR-linear, SVR-RBF, dan SVR kernel polynomial.

Tabel 4.9 Hasil optimisasi hyperparameter menggunakan Metode Grid Search

Metode	alpha	Gamma	epsilon	C	degree
Kernel Ridge	0.01	0.01			
Linear Regression					
SVR-linear		0.1	0.1	4	
SVR-RBF		0.1	0.1	47	
SVR-Poly		0.1	0.1	34	1

Dari data dalam Tabel 4.9, Hasil optimisasi hyperparameter dari fungsi Grid Search menggunakan python, ditunjukkan di dalam tabel tersebut bahwa Kernel Ridge Regression hanya menggunakan dua hyperparameter yaitu alpha dan gamma. Nilai gamma sama dengan 0.01 berarti parameter kernel rbf sebesar 0.01 dan nilai alpha 0.01. Di baris metode Linear Regression nilainya kosong karena metode ini tidak mempunyai parameter yang perlu dioptimisasi. Metode SVR-rbf, SVR-Linear dan SVR-Poly menggunakan hyperparameter yang sama yaitu gamma, epsilon dan C. Walau data masukan dari SVR sama, dan range parameter C, alpha, gamma dan epsilon sama tetapi nilai C yang dihasilkan antara kernel tidak sama. Khusus untuk SVR-Poly menggunakan hyperparameter tambahan yaitu degree yang berarti derajat polinomial dari SVR. Hyperparameter tersebut digunakan untuk proses training / fitting data sehingga menghasilkan

sebuah model prediksi yang dapat digunakan untuk mengestimasi nilai IHK harian.

Jika menggunakan metode optimisasi hyperparameter Random Search, menghasilkan nilai score yang bernilai koefisien determinant / R^2 seperti yang digunakan metode Grid Search. Data hasil dari metode Random Search untuk medel SVR rbf ditunjukkan di dalam Tabel 4.10.

Tabel 4.10. Hasil dari metode Random Search untuk SVR-RBF

Percobaan	score	C	gamma	epsilon
1	0.99	36	0.1	0.1
2	0.99	44	0.1	0.1
3	0.99	45	0.1	0.1
4	0.99	46	0.1	0.1
5	0.99	47	0.1	0.1
6	0.99	48	0.1	0.1
7	0.99	49	0.1	0.1
8	0.99	50	0.1	0.1
9	0.99	52	0.1	0.1
10	0.99	53	0.1	0.1
11	0.99	58	0.1	0.1

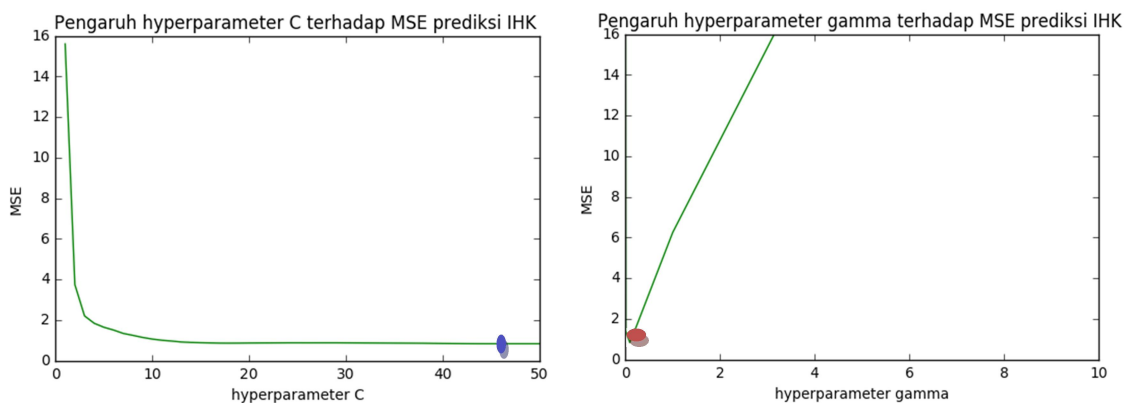
Di dalam Tabel 4.10, kolom percobaan menunjukkan langkah percobaan yang dilakukan. Setiap satu kali percobaan, metode Random Search dijalankan, dan menghasilkan 1 nilai score dan 1 kombinasi parameter terbaik C, gamma dan epsilon. Dapat dilihat bahwa score yang dihasilkan setiap percobaan mempunyai nilai yang sama. Nilai hyperparameter gamma dan epsilon besarnya sama. Berbeda dengan hyperparameter C. Nilai C tidaklah unik. Hyperparameter C yang dihasilkan oleh metode Random Search berubah ubah walaupun nilai score yang dihasilkan tetap sama dengan 0.99. Di dalam metode Grid Search yang telah dibahas, nilai C terbaiknya adalah sama dengan 47, pada metode Random Search nilai C sebesar 47 juga termasuk nilai hyperparameter C terbaik.

Pengaruh hyperparameter terhadap nilai evaluasi MSE dalam prediksi IHK

Jika optimisasi hyperparameter dilakukan secara manual atau dicoba menentukan semua kombinasi variable tanpa menggunakan metode Grid Search dan Random

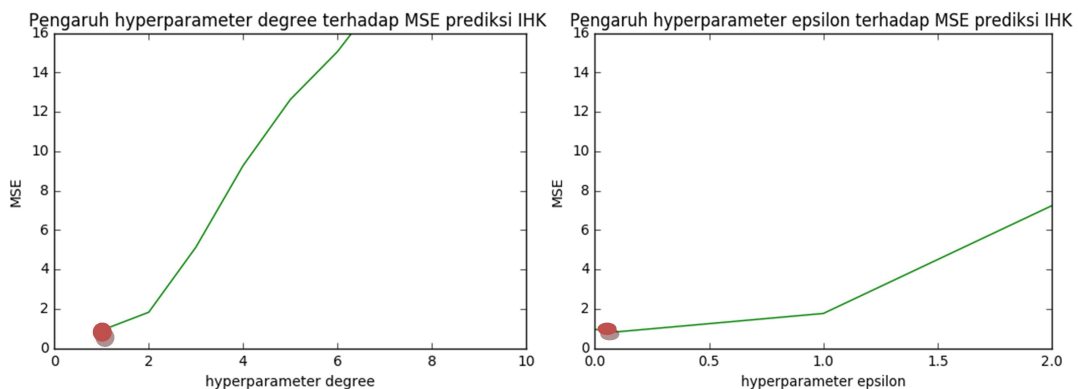
Search, hal ini akan berjalan sangat lama dan rentan kekeliruan. Oleh karena itu dalam penelitian ini menggunakan metode optimisasi Grid Search dan Random Search untuk mencari parameter yang optimal. Selain nilai hyperparameter tersebut berpengaruh terhadap nilai evaluasi koefisien determinant atau R^2 , nilai hyperparameter ini berpengaruh juga di dalam nilai evaluasi mse.

Didalam Gambar 4.1 ditunjukkan, bagaimana pengaruh hyperparameter C dan gamma terhadap nilai evaluasi model prediksi IHK menggunakan MSE.



Gambar 4.1 Pengaruh hyperparameter C dan gamma terhadap nilai MSE

Sedangkan pengaruh hyperparameter degree dan epsilon ditunjukkan oleh Gambar 4.2. Titik berwarna merah di dalam Gambar 4.1 dan Gambar 4.2 merupakan titik optimal nilai hyperparameter yang diperoleh dengan metode Grid Search.



Gambar 4.1 Pengaruh hyperparameter epsilon dan degree terhadap nilai MSE

Makin banyak nilai C, maka prediksi IHK berada di kondisi konvergen atau MSE nya lebih kecil. Hasil dari mse untuk pengujian dataset menggunakan metode SVR-RBF yang dihasilkan ditunjukkan di dalam Tabel 4.11.

Tabel 4.11 Hasil nilai mse setiap nilai hyperparameter C

C	mse	C	mse	C	mse	C	mse	C	mse
1	15.5986	21	0.8654	41	0.8346	61	0.8050	81	0.7855
2	3.7422	22	0.8674	42	0.8313	62	0.8022	82	0.7852
3	2.1975	23	0.8696	43	0.8288	63	0.7996	83	0.7851
4	1.8336	24	0.8721	44	0.8274	64	0.7973	84	0.7850
5	1.6376	25	0.8724	45	0.8270	65	0.7948	85	0.7848
6	1.4999	26	0.8715	46	0.8268	66	0.7925	86	0.7848
7	1.3332	27	0.8720	47	0.8274	67	0.7903	87	0.7846
8	1.2375	28	0.8727	48	0.8270	68	0.7888	88	0.7846
9	1.1354	29	0.8727	49	0.8268	69	0.7886	89	0.7846
10	1.0545	30	0.8701	50	0.8267	70	0.7882	90	0.7846
11	0.9967	31	0.8677	51	0.8268	71	0.7878	91	0.7844
12	0.9579	32	0.8650	52	0.8266	72	0.7877	92	0.7845
13	0.9124	33	0.8627	53	0.8268	73	0.7874	93	0.7844
14	0.8895	34	0.8603	54	0.8243	74	0.7869	94	0.7844
15	0.8769	35	0.8580	55	0.8221	75	0.7868	95	0.7843
16	0.8666	36	0.8558	56	0.8192	76	0.7865	96	0.7842
17	0.8573	37	0.8533	57	0.8164	77	0.7863	97	0.7842
18	0.8553	38	0.8497	58	0.8133	78	0.7861	98	0.7844
19	0.8581	39	0.8436	59	0.8106	79	0.7859		
20	0.8628	40	0.8383	60	0.8076	80	0.7856		

Data di dalam Tabel 4.11 menunjukkan bahwa makin besar nilai C yang digunakan di dalam model SVR-rbf maka nilai mse nya semakin kecil. Perlu diingat, seperti yang telah di bahas di dalam subbab 2.3 tentang Support Vector Regression, bahwa C adalah parameter penalty untuk mengatasi ketidaklayakan pembatas. Nilai C yang bernilai terlalu besar, bisa jadi membuat nilai evaluasi mse menjadi lebih kecil, tetapi permasalahan overfitting harus di hindari. Dimana untuk data ini nilai MSE bernilai bagus, tetapi tidak berlaku untuk data yang baru.

Hasil optimisasi SVR kernel RBF menggunakan metode Grid Search yaitu $C = 47$, gamma 0.1 dan epsilon 0.1. Nilai C sebesar 47, ditunjukkan pada Gambar 4.1 sumbu x dalam grafik. Nilai C ini menghasilkan nilai MSE model

prediksi rendah. Begitu juga nilai gamma dan epsilon sebesar 0.1. Di dalam grafik menunjukkan bahwa pada titik merah, nilai gamma dan epsilon sebesar 0.1, mempunyai nilai MSE prediksi paling rendah dibandingkan nilai gamma dan epsilon lainnya. Hyperparameter degree hanya digunakan pada SVR kernel poly dimana degree yang optimal sebesar 1, ditunjukkan di dalam Gambar 4.2 gambar pengaruh degree. Nilai degree optimal bernilai 1 sama dengan nilai degree hasil dari metode optimisasi dari Grid Search. Dibandingkan dengan nilai hyperparameter C hasil dari metode optimisasi Random Search, nilai C sebesar 36, 44, 45, 46, 47, 48, 49, 50, 52, 53, 58 bernilai optimal. Di dalam Gambar 4.1, nilai C dari Random Search juga menghasilkan MSE yang rendah.

Ini berarti nilai hyperparameter C di dalam model prediksi IHK tidaklah unik, karena bisa menggunakan nilai 36, 36, 44, 45, 46, 47, 48, 49, 50, 52, 53, 58 sedangkan nilai dari hyperparameter gamma, epsilon dan degree merupakan nilai yang unik karena metode Grid Search dan Random search menunjukkan hasil optimisasi yang sama yaitu nilai gamma optimal sebesar 0.1 dan nilai epsilon optimal sebesar 0.1. Keadaan nilai optimal gamma dan epsilon yang unik ini juga ditunjukkan di dalam Gambar 4.1 dan 4.2 dimana nilai mse terendah yang diperoleh model jika menggunakan nilai gamma optimal 0.1 dan nilai epsilon 0.1.

Dengan menggunakan hyperparameter optimal yang telah diperoleh maka hasil evaluasi dari prediksi IHK menggunakan data bulanan ditunjukkan di dalam Tabel 4-12.

Tabel 4.12 Hasil evaluasi pembangunan prediksi IHK bulanan

Metode	MSE Train	R^2 Train	MSE Test	R^2 Test
Kernel Ridge	0.1086	0.9992	0.703	0.9984
linear Regression	0.078	0.9985	1.2364	0.9902
SVR-linear	0.1634	0.9978	0.6833	0.9946
SVR-RBF	0.0557	0.9992	0.2065	0.9984
SVR-Poly	0.2084	0.9971	0.6968	0.9945


Dari Tabel 4.12 dapat dilihat hasil prediksi IHK bulanan menggunakan metode Kernel Ridge dan Linear regression. Walaupun sama-sama dari model linear tetapi keduanya mempunyai hasil yang berbeda. Kernel Ridge Regression mempunyai MSE training sebesar 0.1086 sedangkan Linear Regression mempunyai MSE yang sedikit lebih baik yaitu sebesar 0.078. Namun, pada saat dilakukan pengujian data testing, MSE data testing dari metode Kernel Ridge Regression sebesar 0.703, nilai MSEnya lebih baik daripada Linear Regression yang mempunyai MSE tertinggi diantara metode lainnya yaitu sebesar 1.2364. Artinya, Kernel Ridge Regression mempunyai akurasi prediksi data atau MSE yang lebih baik daripada Linear regression.

Dari semua metode yang dicoba, MSE yang terkecil untuk data test dan data training adalah metode SVR dengan kernel RBF yaitu MSE untuk data train sebesar 0.0557 dan MSE untuk data testing sebesar 0.2065

Model prediksi bulanan digunakan untuk mengestimasi nilai IHK harian. Algoritma estimasi nilai IHK, seperti yang dijelaskan di dalam subbab 3.6.2, dijalankan. Hasil dari estimasi nilai IHK harian menghasilkan data yang ditunjukkan di dalam Table 4-13.

Tabel 4.13 Selisih antara nilai aktual IHK dengan rata bulanan IHK prediksi

IHK bulan Jan 2012, Feb 2012, Dec 2016						
Bulan	Rata-rata IHK prediksi tiap bulanan					IHK- aktual
	KRR	LR	Slin	Srbf	Spoly	
Jan-12	99.19	98.80	99.66	98.97	99.82	98.2
Feb-12	98.82	97.82	99.17	99.05	99.29	98.55
Dec-16	124.90	124.04	123.95	125.39	123.91	125.77



Selisih antara IHK hasil prediksi dengan nilai aktual IHK					
Bulan	KRR	LR	Slin	Srbf	Spoly
Jan-12	0.99	0.60	1.46	0.77	1.62
Feb-12	0.27	0.73	0.62	0.50	0.74
Dec-16	0.87	1.73	1.82	0.38	1.86

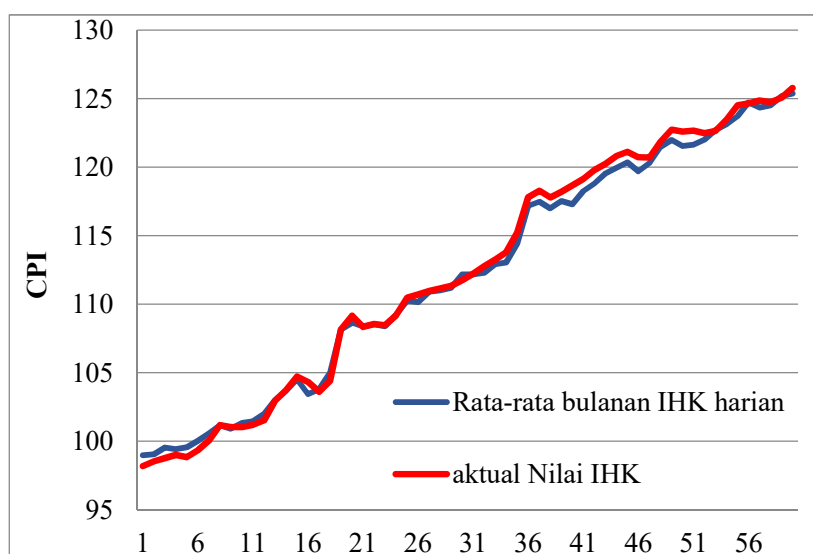
Table 4.13, tabel selisih antara nilai aktual dengan nilai prediksi, ini menunjukkan bagaimana hasil pemilihan IHK yang baik untuk bulan Januari 2012, Februari 2012 dan December 2016.

Tabel 4.14 Performansi IHK hasil prediksi terhadap IHK aktual

Method	MSE	R ²
Kernel Ridge Regression	0.7423	0.9906
Linear Regression	0.7498	0.9905
SVR-Linear	1.0436	0.9867
SVR-rbf	0.3454	0.9956
SVR-Poly	1.0849	0.9862

Dari data di dalam Tabel 4.14, performansi IHK aktual dan prediksi tersebut disimpulkan bahwa metode SVR dengan menggunakan kernel rbf mempunyai MSE lebih kecil sebesar 0.3454 dibandingkan metode Kernel Ridge sebesar 0.7423, Linear Regression sebesar 0.7498, SVR-linear 1.0436, SVR Poly 1.0849. Jika dihubungkan dengan standard Deviasi yang ditetapkan oleh Bank Indonesia, nilai standard deviasi yang ditoleransi di dalam prediksi sebesar 1, maka hasil mse sebesar 0.343 dan MSE berupa kuadrat dari standard deviasi maka nilai mse ini masih di dalam nilai standard deviasi yang ditoleransi.

Hasil prediksi IHK diatas ditunjukkan ke dalam bentuk Gambar 4-3, yaitu grafik perbandingan rata-rata bulanan IHK harian menggunakan metode SVR kernel rbf dengan nilai aktual IHK. Sumbu x dalam grafik menunjukkan bulan ke-n sedangkan sumbu y menunjukkan nilai IHK. Garis berwarna biru di dalam gambar menunjukkan grafik rata-rata bulanan dari IHK hasil prediksi. Sedangkan garis berwarna merah menunjukkan IHK aktual dalam frekuensi bulanan. Dari gambar itu dapat disimpulkan metode SVR menggunakan kernel RBF mampu memprediksi IHK dengan menggunakan data harian. Terlihat perbedaan garis IHK hasil prediksi yang mendekati nilai aktual IHK berarti nilai IHK hasil prediksi mendekati nilai aktualnya.



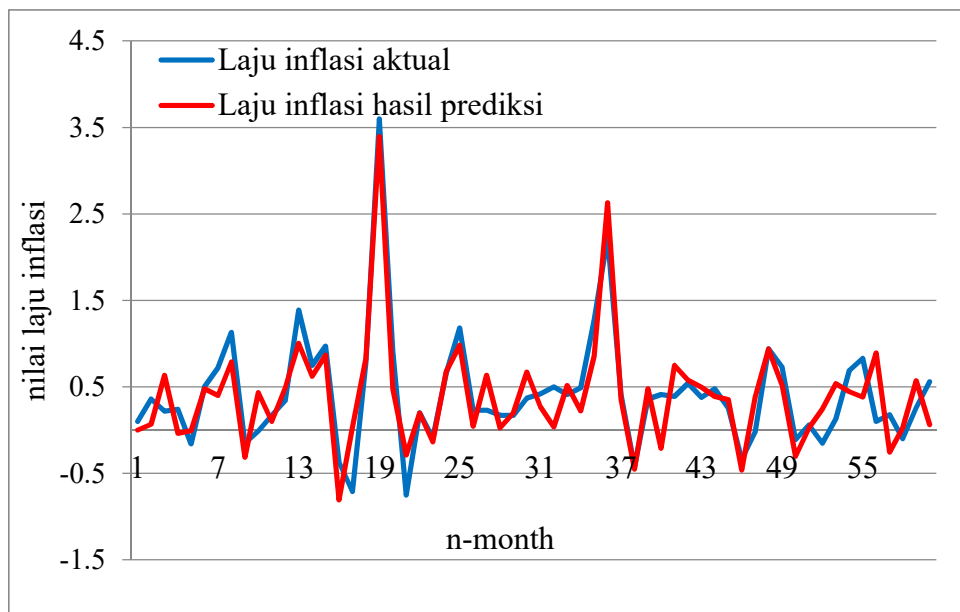
Gambar 4.2 Grafik perbandingan rata-rata bulanan IHK harian menggunakan metode SVR kernel rbf dengan nilai aktual IHK

Hasil prediksi IHK harian tersebut dalam mengenali titik IHK yang mengalami kenaikan laju inflasi.

Dalam Tabel 4.6 telah ditunjukkan perubahan IHK dengan basis tahun yang sama di kota Surabaya. Didalam data tersebut, nilai laju inflasi terdapat didalam kolom nilai inflasi. Nilai aktual laju inflasi yang diatas 1% dibandingkan dengan nilai laju inflasi hasil prediksi. Data tersebut ditunjukkan di Tabel 4.15, table perbandingan laju inflasi aktual dan prediksi dan Gambar 4.4.

Tabel 4.15 Perbandingan laju inflasi aktual dan prediksi

bulan	IHK_bulanan	laju-inflasi	IHK-prediksi	Laju-inflasi prediksi
Aug-12	101.19	1.13	101.23	0.79
Jan-13	102.95	1.39	102.97	1.01
Jul-13	108.17	3.6	108.11	3.39
Jan-14	110.47	1.18	110.19	0.98
Nov-14	115.24	1.27	114.08	0.86
Dec-14	117.81	2.23	117.08	2.63



Gambar 4.3 Perbandingan Laju inflasi aktual dan hasil prediksi

Dari Gambar 4.4, gambar Perbandingan Laju inflasi aktu dan hasil prediksi dapat dilihat perbandingan laju inflasi aktual dan hasil prediksi. Hasil yang diperoleh, selisih antara laju inflasi aktual dan hasil prediksi dibawah 1 %. Model prediksi ini mampu mengenali pola saat laju inflasi mengalami kenaikan karena mempunyai selisih dibawah 1%.

Hasil dari prediksi IHK harian tersebut menjadi dataset harian dengan besar data 1827 data. Dataset harian yang diperoleh berisi variabel masukan X, nilai komoditas dan kurs rupiah terhadap dollar dengan frekuensi harian, dan variabel targetnya, y berisi nilai IHK harian hasil prediksi.

4.4 Hasil pembangunan model prediksi IHK harian

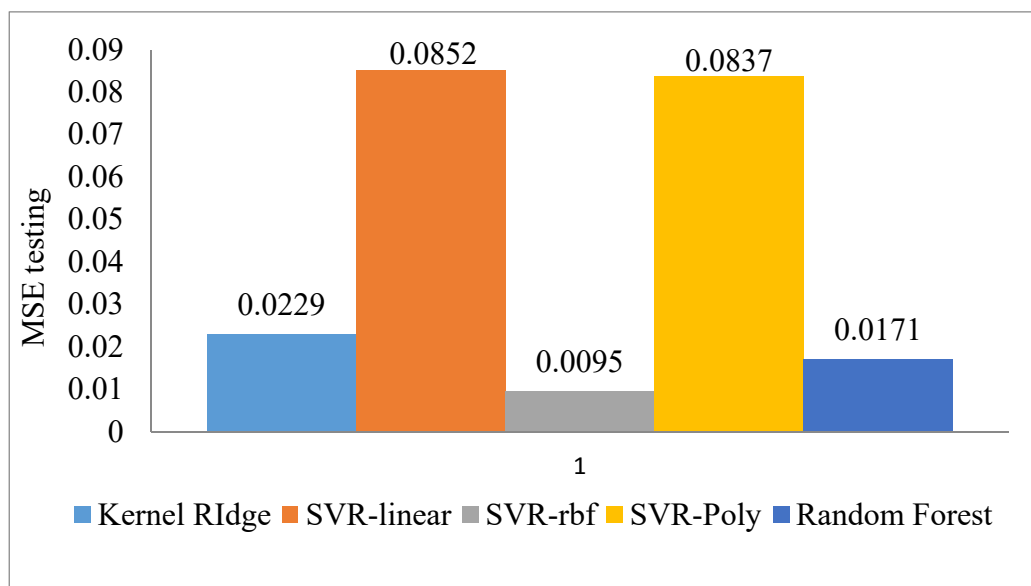
Prediksi IHK menggunakan data harian ini merupakan perbaikan dari model prediksi bulanan. Setelah memperoleh estimasi nilai IHK harian maka prediksi harian dibangun menggunakan metode Kernel Ridge Regression, SVR, Random Forest. Metode Linear regression tidak digunakan didalam pembangunan model harian karena mse nya rendah sebesar 0.7498 . Yang difokuskan dalam penelitian tahap ini adalah akurasi model IHK harian dan perhitungan waktu yang dibutuhkan dalam training dan testing data. Hasil pembangunan prediksi harian dengan lima metode, Ridge Regression, SVR-linear, SVR-rbf, SVR-poly dan

Random Forest ditunjukkan dalam Tabel 4.16 yaitu table nilai MSE dan waktu komputasi dalam prediksi IHK harian.

Tabel 4.16 Nilai MSE dan waktu komputasi prediksi IHK harian

Metode	MSE-Train	MSE-Test	Waktu Training	Waktu Testing
KernelRidge	0.0191	0.0229	0.0698	0.0198
SVR-linear	0.0708	0.0852	2.1764	0.0132
SVR-RBF	0.0050	0.0095	0.134	0.0121
SVR-Poly	0.0702	0.0837	0.2923	0.0152
Random Forest	0.0061	0.0171	0.1205	0.0030

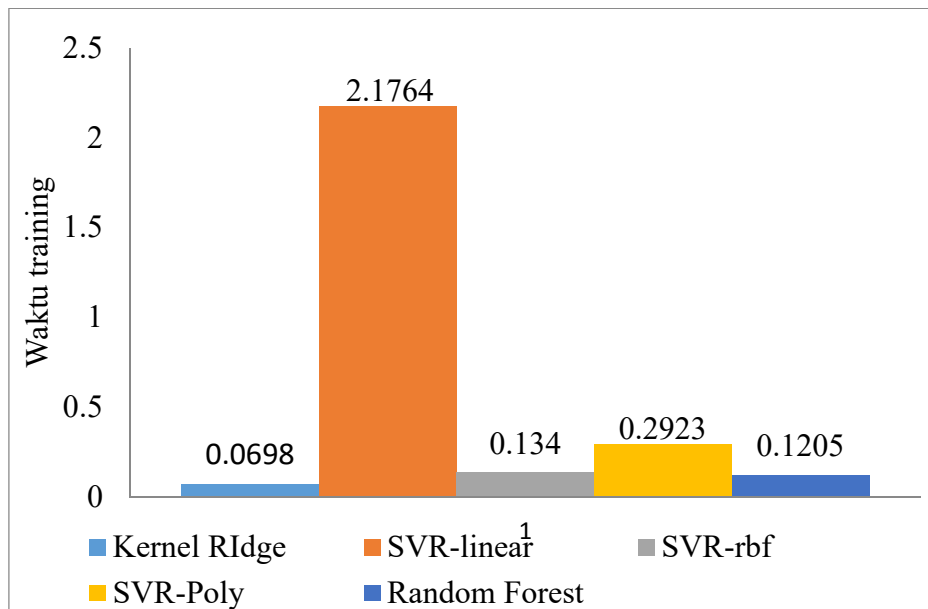
Dari percobaan yang dilakukan, dari data di dalam Tabel 4.16 diambil kesimpulan bahwa SVR linear mempunyai waktu training lebih besar yaitu sebesar 2.1764 detik. Sedangkan MSE paling kecil sebesar 0.0050 pada SVR dengan kernel RBF.



Gambar 4.4 Perbandingan nilai MSE yang diperoleh saat testing prediksi IHK dengan lima metode prediksi

Dari sisi akurasi prediksi, SVR-kernel RBF memiliki MSE paling baik untuk training sebesar 0.0050 dan mse testing data sebesar 0.0095, diikuti dengan

metode random forest yang mempunyai selisih yang cukup sedikit dengan SVR-kernel RBF yaitu mse training sebesar 0.0061 dan mse testing sebesar 0.00171.



Gambar 4.5 Perbandingan waktu training prediksi IHK harian

MSE yang diperoleh didalam proses training dan testing semua metode prediksi kurang dari nilai 0.01. Di dalam Gambar 4.6 ditunjukkan bahwa dari sisi waktu, waktu yang dibutuhkan untuk training, paling lama di metode SVR Linear sebesar 2.1764, sedangkan yang paling cepat Kernel Ridge sebesar 0.0698, posisi kedua Random forest sebesar 0.1205 dan SVR kernel RBF sebesar 0.134. SVR – kernel RBF dan Random Forest bisa digunakan untuk memprediksi data harian karena mse nya yang lebih kecil atau kurang dari 1 dan waktu komputasinya lebih sedikit.

Hasil Prediksi IHK harian dalam mengenali data harian yang baru

Mengambil contoh Data Kota Surabaya pada bulan Januari 2017 dimana :

IHK aktual pada bulan Desember 2016 sebesar 125.77.

IHK aktual Januari 2017 sebesar 127.98. Hasil prediksi ditunjukkan oleh Tabel 4.17 yaitu tabel data hasil prediksi IHK harian Kota Surabaya pada bulan Januari 2017. Rata – rata IHK prediksi pada bulan Januari 2017 sebesar 126.89 selisih

antara nilai IHK aktual sebesar 127.98 dengan hasil prediksi sebesar 126.89 masih tinggi yaitu sebesar 1.09.

Tabel 4.17 Data Hasil Prediksi Bulan Januari 2017

Tanggal	IHK	Tanggal	IHK
31 Des 2016	125.09	16 Jan 2017	127.20
1 Jan 2017	125.40	17 Jan 2017	127.23
2 Jan 2017	125.40	18 Jan 2017	126.90
3 Jan 2017	126.90	19 Jan 2017	127.10
4 Jan 2017	126.95	20 Jan 2017	127.08
5 Jan 2017	127.24	21 Jan 2017	127.10
6 Jan 2017	126.92	22 Jan 2017	127.08
7 Jan 2017	126.90	23 Jan 2017	126.96
8 Jan 2017	126.90	24 Jan 2017	126.90
9 Jan 2017	126.90	25 Jan 2017	126.90
10 Jan 2017	126.90	26 Jan 2017	126.95
11 Jan 2017	127.14	27 Jan 2017	126.93
12 Jan 2017	126.99	28 Jan 2017	126.92
13 Jan 2017	127.07	29 Jan 2017	126.90
14 Jan 2017	127.09	30 Jan 2017	126.90
15 Jan 2017	127.10	31 Jan 2017	126.75

Tetapi jika melihat lebih detil ke waktu, kapan laju inflasi tersebut mulai tinggi, diperoleh hasil bahwa mulai tanggal 3 Januari selisih IHK sudah mulai tinggi, nilai IHK tanggal 2 Januari sebesar 125.4 sedangkan nilai IHK tanggal 3 Januari sebesar 126.9. Dengan prediksi harian ini bisa dilihat pola data harian, sehingga prediksi IHK bisa diketahui lebih cepat. Kenaikan laju inflasi sudah mulai bisa diketahui sejak tanggal 3 Januari 2017. Gambar 4.7 menggambarkan kondisi IHK pada bulan Januari 2017, terlihat kenaikan yang cukup signifikan dari tanggal 1 sampai dengan 5 Januari 2017



Gambar 4.6 nilai IHK prediksi pada bulan Januari 2017

4.5 Waktu komputasi dalam optimisasi hyperparameter

Berdasarkan hasil percobaan yang dilakukan, di dalam Tabel 4.18 ditunjukkan bahwa metode Grid Search menggunakan waktu komputasi lebih lama yaitu sebesar 700.07 detik untuk 1461 data dengan menggunakan metode SVR-rbf dibandingkan waktu komputasi Random Search sebesar 74.52 detik. Metode ini menghasilkan sebuah nilai yang dikenal dengan best score. Best score yang dihasilkan adalah sama dengan nilai *koefisien determinan* atau R^2 . Antara kedua metode optimisasi hyperparameter, R^2 terbaiknya tidak banyak perbedaan, untuk data 1461 menggunakan metode SVR-rbf mempunyai best score sebesar satu begitu juga dengan Random search yang mempunyai best score sebesar satu.

Perbedaan waktu komputasi yang dibutuhkan antara Grid Search dan Random Search untuk lima metode prediksi di tunjukkan di dalam Gambar 4.7, di dalam data bervolume 1278, 1461, 1644 terlihat perbedaan yang jelas garis antara Grid Search dan garis Random mempunyai perbedaan yang jauh untuk 4 metode prediksi yaitu SVR kernel Linear, SVR- rbf, SVR- poly dan Random Forest.

Tabel 4.18 Waktu komputasi dan best score dari metode optimisasi

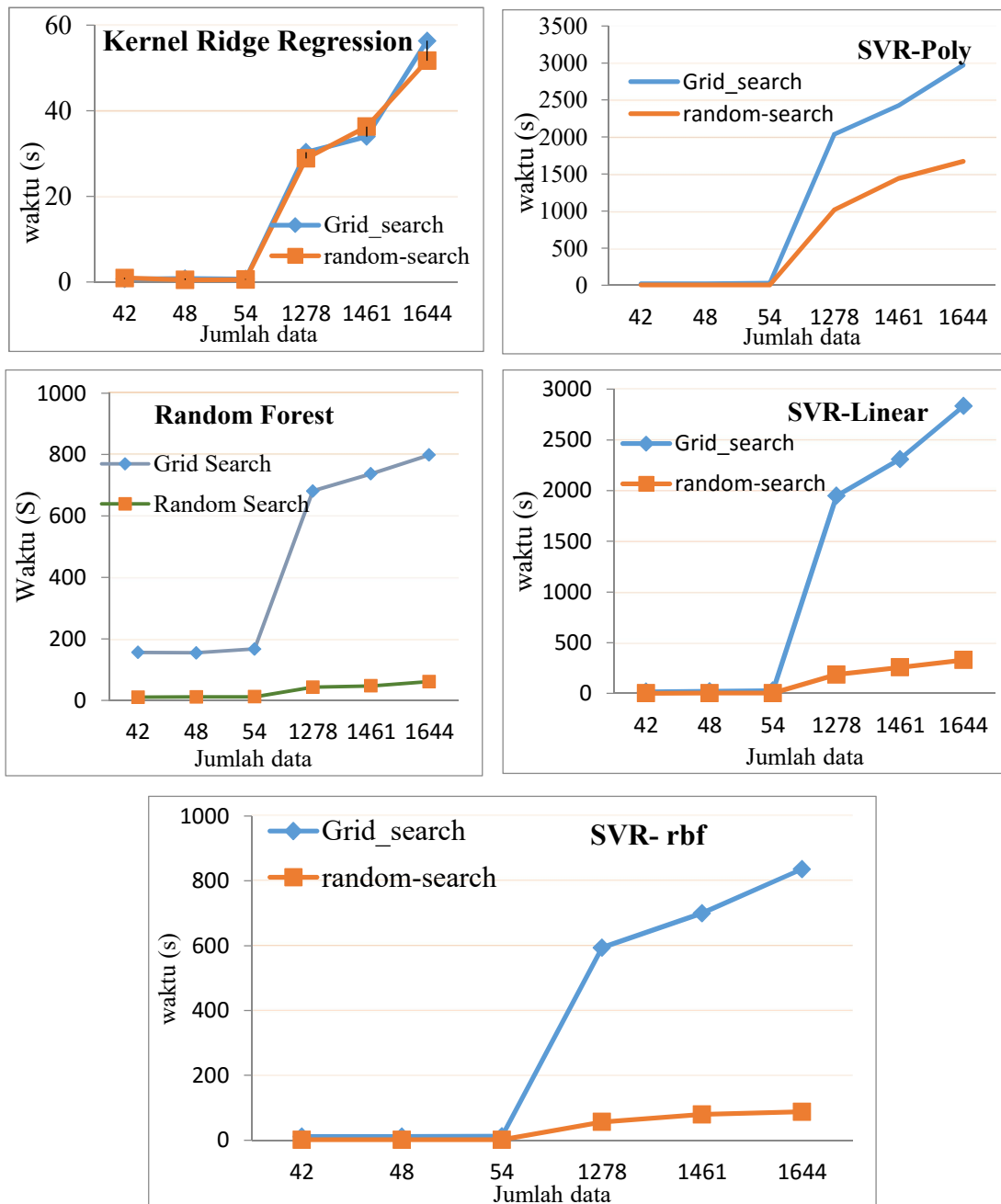
Metode	n	PC			
		Grid	R ²	Random	R ²
Ridge Regression	42	0.70	0.99	0.93	0.98
SVR-linear		15.97	0.99	1.87	0.99
SVR-RBF		11.50	0.99	1.47	0.99
SVR-Poly		20.08	0.99	1.88	0.99
RFR		156.02	0.98	10.04	0.98
Ridge Regression	48	0.78	0.99	0.49	0.97
SVR-linear		20.35	0.99	2.41	0.97
SVR-RBF		11.64	0.99	1.35	0.99
SVR-Poly		20.54	0.99	2.41	0.97
RFR		154.47	0.99	10.86	0.99
Ridge Regression	54	0.68	0.99	0.55	0.98
SVR-linear		27.60	0.99	3.08	0.99
SVR-RBF		12.42	0.99	1.59	0.99
SVR-Poly		27.34	0.99	2.85	0.99
RFR		167.03	0.99	11.63	0.99
Ridge Regression	1278	30.34	1.00	28.86	1.00
SVR-linear		1947.97	1.00	186.12	1.00
SVR-RBF		594.05	1.00	56.38	1.00
SVR-Poly		2038.74	1.00	1018.55	1.00
RFR		681.24	1.00	42.37	1.00
Ridge Regression	1461	32.82	1.00	26.65	1.00
SVR-linear		2309.57	1.00	293.65	1.00
SVR-RBF		700.07	1.00	74.52	1.00
SVR-Poly		2428.03	1.00	546.56	1.00
RFR		737.21	1.00	46.83	1.00
Ridge Regression	1644	38.73	1.00	51.71	1.00
SVR-linear		2833.64	1.00	328.86	1.00
SVR-RBF		836.54	1.00	87.79	1.00
SVR-Poly		2970.81	1.00	1672.87	1.00
RFR		798.49	1.00	60.94	1.00

Hasil dari perbandingan metode Grid Search dan Random Search dalam mencari model yang terbaik ditunjukkan pada Gambar 4.7, gambar performansi metode Optimisasi hyperparameter. Dalam grafik menunjukkan Kernel Ridge Regression mempunyai waktu yang nyaris sama antara dua metode optimisasi, ini

bisa dikarenakan, range hyperparameter dari kernel Ridge Regression hanya dua sehingga range variabel Ridge regression yang diobservasi juga sama.

Penggunaan waktu tertinggi adalah waktu yang digunakan SVR kernel Polynomial, karena hyperparameter yang perlu dioptimisasi lebih banyak dibandingkan metode lainnya yaitu C, gamma, epsilon dan degree. Urutan kedua yang lebih lama adalah SVR kernel Linear. Data di dalam Tabel 4.18 tersebut menunjukkan nilai R^2 dan waktu eksekusi keduanya. Jika eksperimen dilakukan dengan waktu terbatas, direkomendasikan penggunaan optimisasi hyperparameter metode Random Search.

Ditunjukkan dalam Tabel 4.18, di dalam model Random Forest, waktu proses untuk optimisasi dengan data 42 sebesar 156.02 detik, 48 sebesar 154.47 detik, 54 sebesar 167.03 lebih tinggi daripada waktu proses metode lainnya misalkan SVR-RBF untuk volume 42 data memerlukan waktu 11.50 detik, untuk 48 data sebesar 11.64, untuk 54 data sebesar 12.42, Random Forest mempunyai perbedaan waktu sekitar 10 kali lipat dari SVR-Rbf tetapi pada saat data yang dilatihkan sebesar lebih dari 1000, waktu komputasi random forest tidak setinggi SVR-Linear dan SVR-poly, bahkan waktu komputasinya sama dengan SVR-rbf. Untuk volume data 1278 metode prediksi Random Forest memerlukan waktu 681.24 detik, untuk 1461 data sebesar 737.21, 1644 sebesar 798.49. Sedangkan waktu yang dibutuhkan metode prediksi SVR-rbf untuk 1278 data sebesar 594.05, untuk 1461 data sebesar 700.07 dan untuk 1644 data sebesar 836.54. Dapat di ambil kesimpulan, untuk data yang besar, Random Forest memiliki waktu komputasi yang cukup rendah



Gambar 4.7 Performansi metode Optimisasi hyperparameter

4.6 Hasil Evaluasi setelah migrasi prediksi ke Cloud

4.6.1 Hasil evaluasi prosedur dan fungsionalitas

Disini akan dibahas bagaimana pelaksanaan proses migrasi model prediksi IHK dari personal ke Cloud Computing. Di Subbab 3.9 telah dibahas

proses perencanaan dari framework Cloud-RMM, dalam bab ini dibahas bagaimana proses Pelaksanaan dan Evaluasi migrasi prediksi IHK ke dalam lingkungan Cloud.

Dalam tahapan migrasi, semua prosedur migrasi prediksi IHK harian dari personal komputer ke dalam cloud telah dilakukan. Hasil evaluasi prosedur dan fungsionalitas dirangkum di dalam Table 4.19, tabel evaluasi prosedur dan fungsionalitas.

Tabel 4.19 evaluasi prosedur dan fungsionalitas

No	Fungsi	PC	AWS	Azure
1	Instalasi Virtual Machine	-	Berhasil	Berhasil
2	Instalasi Python	Berhasil	Berhasil	-
3	Instalasi Jupyter notebook	Berhasil	Berhasil	-
4	Akses SSH ke virtual Machine	-	Berhasil	Berhasil
5	Akses Jupyter Notebook melalui browser	Berhasil	Berhasil	Gagal
6	Migrasi File Script dan dataset	Berhasil	Berhasil	Gagal
7	Akses dataset dari jupyter notebook	Berhasil	Berhasil	Gagal

Pembahasan dari migrasi tersebut, dibedakan berdasarkan nama penyedia layanan cloud berikut :

1. Amazon Web Service(AWS)

Instalasi *Virtual machine* atau VM dilakukan di layanan EC2. Paket instalasi yang disediakan AWS tidak berisi paket *python* sehingga *python* dan *jupyter notebook* harus diinstal manual. Semua prosedur migrasi yang telah digambarkan di dalam Gambar 3-15, alur migrasi model prediksi ke cloud computing, berhasil diterapkan tanpa ada kendala prosedur. Tidak dilakukan modifikasi kode pemrograman aplikasi dan ekstraksi data dilakukan di harddisk dari *virtual machine*.

2. Microsoft Azure

Microsoft Azure menyediakan paket instalasi data analytic. Sehingga pengguna data *analytic* tidak perlu menginstall python dan Jupyter notebook. Instalasi python dan *jupyter notebook* dapat dijalankan. Akses SSH melalui terminal console bisa dilakukan. Tetapi akses *Jupyter Notebook* melalui browser tidak bisa dilakukan. Di dalam prosedur yang telah digambarkan di bab 3.9 telah dijelaskan, cara alternatif jika *jupyter Notebook* gagal diakses, maka harus mencoba *tunneling* SSH. Setelah langkah alternatif digunakan, Jupyter bisa diakses via remote.

Migrasi file program berhasil dilakukan tetapi dalam migrasi dataset terjadi masalah. Dataset Azure tidak bisa diakses langsung. Dataset Azure ditempatkan di storage Azure. Untuk mengakses dataset tersebut, Azure mempunyai script khusus atau library python khusus Azure yang harus ditambahkan dalam script program (modifikasi kode pemrograman).

```
1. Import library azureml untuk workspace
from azureml import Workspace

2. Mengisi ID dari workspace
ws = Workspace(workspace_id, authorization_token, endpoint)

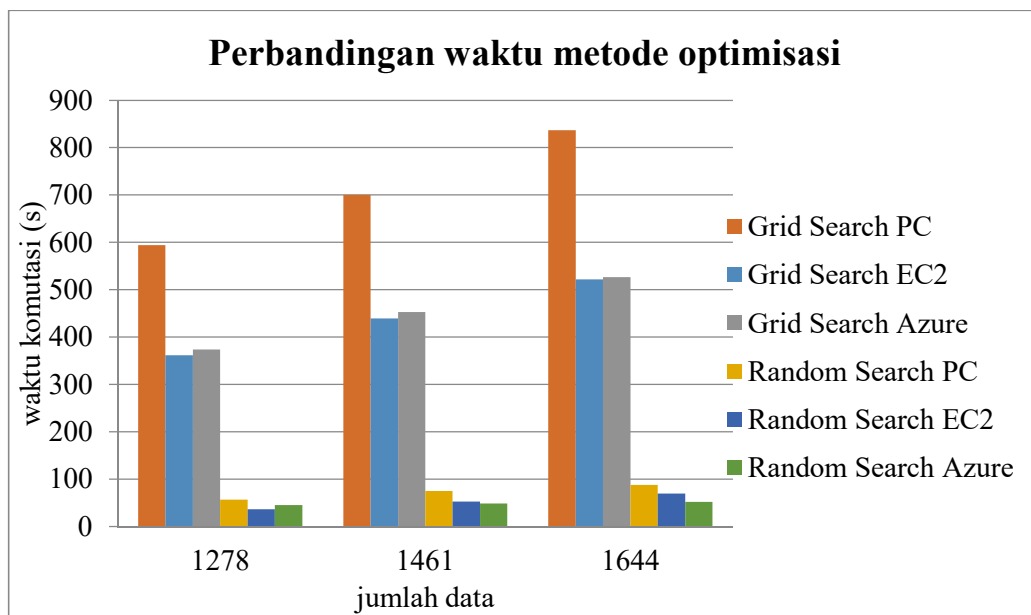
3. Memanggil dataset ke dalam frame data
ds = ws.datasets['monthsurabaya-1704.csv']
```

Setelah proses modifikasi dilakukan, script berhasil mengakses data set.

4.6.2 Hasil evaluasi waktu komputasi optimisasi hyperparameter

Walaupun memory yang digunakan di personal computer lebih besar yaitu 12 GB, dan memory yang disediakan virtual machine di cloud hanya 1GB, jika prediksi IHK dijalankan di Cloud Computing membutuhkan waktu yang lebih sedikit. Terlihat di dalam Gambar 4.9 diagram batang warna orange mewakili PC lebih tinggi daripada batang Cloud computing warna biru dan abu abu. Keadaan yang digambarkan pada Gambar 4.9, untuk waktu komputasi metode SVR- rbf.

Hasil perbandingan waktu optimisasi yang dibutuhkan di dalam EC2 dan di Azure memperlihatkan, waktu proses di dalam EC2 dan Azure, hampir sama. Tabel data tersebut ditunjukkan di dalam Tabel 4.20, hasil evaluasi penggunaan metode Grid dan Random Search untuk lima metode. Makin banyak jumlah data, waktu pemrosesan data makin banyak. Makin banyak hyperparameter maka semakin banyak waktu yang dibutuhkan



Gambar 4.8 Waktu komputasi metode Optimisasi

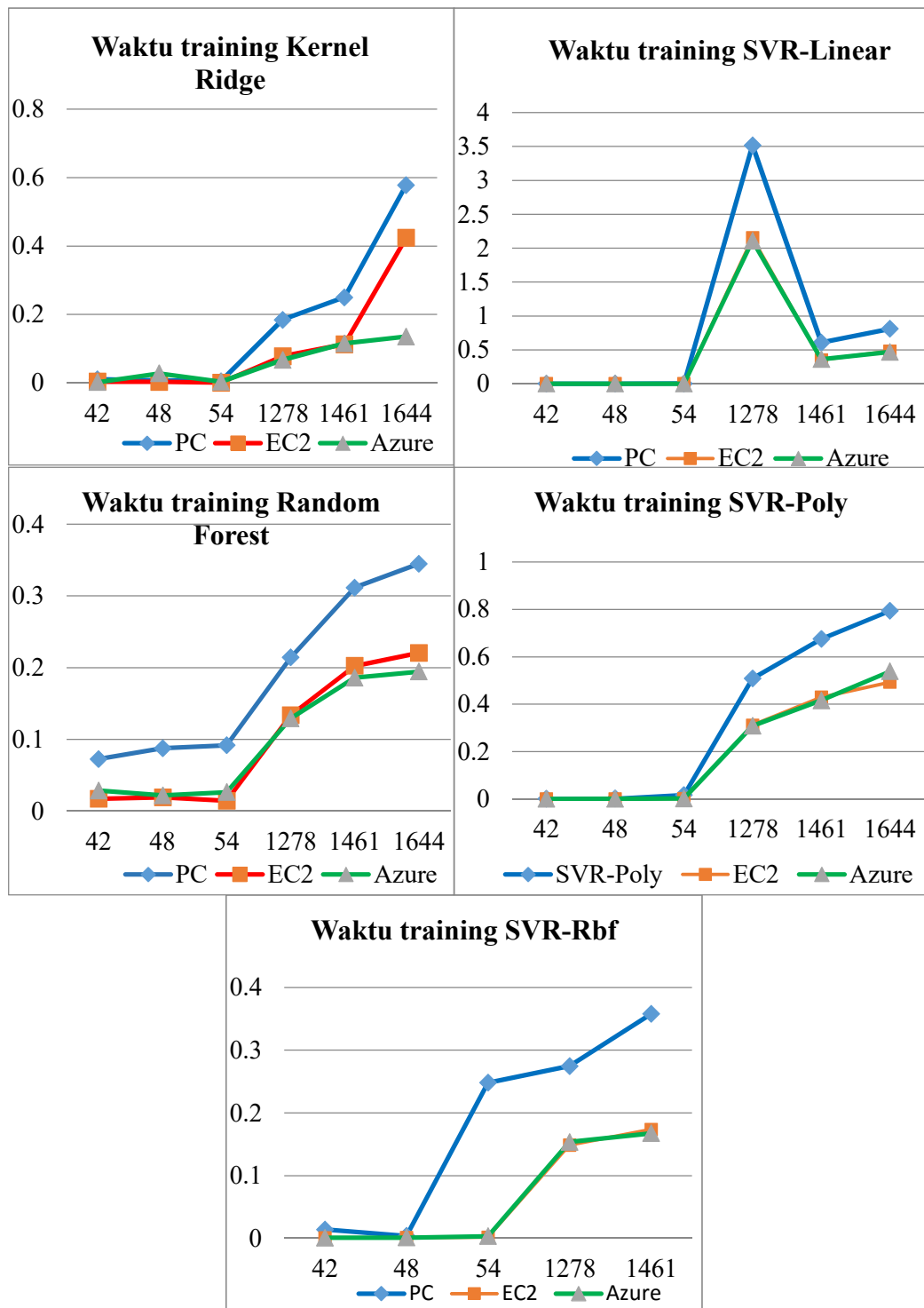
Dari Gambar 4.9 terlihat perbedaan yang cukup signifikan antara Grid search dan Random search di lingkungan komputasi Cloud. Batang dari Grid search untuk PC misal batang warna orange dari PC lebih tinggi daripada batang warna kuning mewakili Random search dari PC. Nilai data di dalam Gambar 4.9 dapat dilihat di Tabel 4.20. Diambil kesimpulan waktu proses untuk Grid Search lebih lama dibandingkan Random Search. Lima metode prediksi, Kernel Ridge, SVR-linear, SVR-Rbf, SVR Poly dan Random Forest dibandingkan pada tiap sumber komputasi.

Tabel 4.20 Hasil evaluasi penggunaan metode Grid dan Random Search

PC	Grid	R²	Random	R²
Ridge Regression	38.734	0.999	51.714	0.999
SVR-linear	2833.641	0.997	328.863	0.997
SVR-RBF	836.538	1	87.787	1
SVR-Poly	2970.814	0.997	1672.866	1
RFR	798.492	0.999	60.936	0.999
Amazon EC2	Grid	R²	Random	R²
Ridge Regression	15.776	0.999	16.354	0.999
SVR-linear	1766.56	0.997	180.269	0.997
SVR-RBF	521.763	1	69.21	1
SVR-Poly	1849.29	0.997	750.044	1
RFR	521.748	0.999	34.865	0.999
Azure	Grid	R²	Random	R²
Ridge Regression	14.15	0.999	8.782	0.999
SVR-linear	1875.75	0.997	236.309	0.997
SVR-RBF	526.433	1	51.652	1
SVR-Poly	1949.71	0.997	670.331	0.999
RFR	530.346	0.999	35.313	0.999

4.6.3 Hasil Evaluasi Waktu Proses Training dan Testing data

Dalam tahap evaluasi ini, dilakukan proses training dengan menggunakan hyperparameter hasil proses Grid Search kemudian diukur waktu training dan testing yang dibutuhkan jika dijalankan di Personal Computer, Amazon EC2 dan Azure. Dari Gambar 4.10 dapat dilihat bahwa waktu training di cloud berjalan lebih singkat daripada di personal computer. Di dalam Gambar 4.10 ditunjukkan waktu training dari lima metode. Data lengkap hasil evaluasi ini dilampirkan dalam thesis ini.



Gambar 4.9 Perbandingan waktu training 5 metode prediksi

4.6.4 Hasil evaluasi nilai MSE dan R^2

Evaluasi nilai MSE dan R^2 di tiga lingkungan komputasi; PC, Microsoft Azure, AWS-EC2 untuk menunjukkan konsistensi model jika dijalankan di ketiganya. Ketiga lingkungan komputasi melakukan training dan testing data yang sama dengan hyperparameter model prediksi juga sama. Dan hasil evaluasi ini membuktikan ketiga lingkungan komputasi mempunyai konsisten yang sama. Hasil yang dikeluarkan ketiga sumber sama.

Tabel 4.21 Hasil evaluasi nilai MSE dan R^2

Metode	PC	Amazon EC2	Azure
Ridge Regression	0.0122/0.9998 0.0132 / 0.9998	0.0122/0.9998 0.0132 / 0.9998	0.0122/0.9998 0.0132 / 0.9998
SVR-linear	0.0720/0.9991 0.0778 / 0.9990	0.0720/0.9991 0.0778 / 0.9990	0.0720/0.9991 0.0778 / 0.9990
SVR-RBF	0.0047/0.9999 0.0077 / 0.9999	0.0047/0.9999 0.0077 / 0.9999	0.0047/0.9999 0.0077 / 0.9999
SVR-Poly	0.0720/0.9990 0.0775 / 0.9990	0.0720/0.9990 0.0775 / 0.9990	0.0720/0.9990 0.0775 / 0.9990
RFR	0.0047/0.9999 0.0120 / 0.9998	0.0047/0.9999 0.0120 / 0.9998	0.0047/0.9999 0.0120 / 0.9998

Keterangan :

Di dalam Tabel 4.21 nilai 0.0122/0.9998 | 0.0132 / 0.9998 berarti

MSE train sebesar 0.0122 , nilai koefisien determinant train sebesar 0.9998 dan MSE test sebesar 0.0132, nilai koefisien determinant test sebesar 0.9998

BAB 5

PENUTUP

5.1 Kesimpulan

Tujuan penelitian ini adalah membangun prediksi IHK menggunakan data harian yaitu harga bahan pokok harian dan kurs rupiah terhadap dollar menggunakan Support Vector Regression. Kemudian mengintegrasikan semua sumber data ke dalam lingkungan cloud computing untuk mempercepat pemrosesan data.

Berdasarkan tujuan tersebut, telah dilakukan percobaan beberapa metode untuk mendapatkan prediksi yang baik dari akurasi yang diwakili dengan MSE dan koefisien determinan. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa Prediksi IHK menggunakan data harian dan menggunakan metode Support Vector Regression bisa digunakan untuk mengembangkan prediksi bulanan. Support Vector Regression yang digunakan menggunakan kernel RBF. Hasil dari prediksi IHK adalah nilai MSE sebesar 0.3454 dengan skala data sama dengan 100 dan koefisien determinan sebesar 0.9991. Dari nilai standard deviasi yang diterapkan oleh Bank Indonesia yaitu sebesar 1, maka nilai MSE ini masih dibawah 1 dan masih bisa ditoleransi. Hyperparameter yang digunakan di dalam SVR rbf ini adalah nilai C sama dengan 47, gamma sama dengan 0.1 dan epsilon sama dengan 0.1. Nilai C ini bersifat tidak unik, karena bisa menggunakan nilai C yang lain seperti 36, 44, 45, 46, 47, 48, 49, 50. Sedangkan nilai hyperparameter lainnya, yaitu epsilon, gamma dan degree bersifat unik. Metode Grid search dan Random search menghasilkan nilai yang sama yaitu nilai gamma optimal 0.1, nilai epsilon optimal sebesar 0.1. Dari simulasi data yang dimigrasikan kedalam lingkungan cloud computing menghasilkan kesimpulan, prediksi IHK menggunakan data harian ini mempunyai waktu pemrosesan Grid search dengan metode SVR-RBF di Cloud Amazon untuk 1644 data sebesar 522 detik, lebih singkat waktunya dibandingkan di PC sebesar 837 detik.

5.2 Saran Penelitian Selanjutnya

Penelitian ini merupakan awal dari penelitian yang berkaitan dengan Big Data dan cloud computing. Banyak tantangan yang belum mungkin diselesaikan di dalam penelitian ini karena kendala waktu. Beberapa diantaranya adalah :

1. Variable harian yang mungkin berkaitan dengan inflasi masih sangat banyak, dan bisa diteliti lebih lanjut. Misalkan, pengaruh twitter, pengaruh variable BBM, Tarif dasar listrik, atau variable yang telah digunakan didalam penelitian sebelum ini. Pengembangan penelitian ini dengan menambahkan variable harian lainnya.
2. Penelitian ini bisa dikembangkan dengan meningkatkan akurasi dari model prediksi melalui metode seleksi fitur. Metode seleksi fitur bisa dilakukan untuk memilih variable mana yang paling berpengaruh dengan pergerakan IHK. Penggunaan seleksi fitur didalam prediksi IHK diharapkan dapat mempercepat waktu komputasi prediksi dan meningkatkan akurasinya.
3. Berkaitan dengan cloud, pengembangan yang belum dilakukan disini adalah memodelkan data prediksi IHK di dalam *cloud computing*. Penelitian dikaitkan dengan fungsi *scalability* dari cloud. Bagaimana kondisi yang ideal dapat diperoleh dengan penambahan spesifikasi hardware yang optimal.
4. Pengembangan thesis ini bisa diarahkan ke pararel processing. Pemrosesan yang pararel memungkinkan data diproses dengan lebih cepat. Di dalam python telah disediakan teknik yang mendukung pararel processing. Di dalam cloud telah disediakan teknologi clustering, sehingga proses bisa dipecah dibeberapa mesin.. Pararel processing akan membagi proses ke dalam unit process. Unit process ini bergantung dengan jumlah CPU yang ada di dalam mesin.
5. Penelitian prediksi dengan menggunakan teknologi Spark bisa dicoba. Pengembangan prediksi IHK dengan menggunakan data harian ini bisa dilakukan secara real time. Prosedure training data secara realtime mungkin berbeda dari yang dibahas disini. Teknologi Spark memiliki kemampuan untuk memproses data secara realtime.

DAFTAR PUSTAKA

- [1] Doran Pandapotan Manik and Albarda, "Using Big Data in Statistics Indonesia," in *International Conference on Information Technology System and Innovations (ICITSI)*, Badung Bali, 2015.
- [2] Novi Wulandari, Setiawan, and Imam Safawi Ahmad, "Peramalan Inflasi Kota Surabaya dengan Pendekatan ARIMA, Variasi Kalender dan Intervensi," *Jurnal Sains dan Seni ITS*, vol. 5, 2016.
- [3] Rumana Hossain and Shaukat Ahmed, "Forecast of Inflation in Bangladesh using ANN model," *IJASCSE*, vol. 2 No. 1, July 2013.
- [4] Gour Sundar Mitra Thakur, Rupak Bhattacharyya, and Seema Sarkar Mondal, "Artificial Neural Network Based Model for Forecasting of Inflation in India," in *Fuzzy Information and Engineering*, 2015, pp. 87-100.
- [5] Linyun Zhang and Jinchang Li, "Inflation Forecasting Using Support Vector Regression," in *Fourth International Symposium on Information Science and Engineering*, 2012.
- [6] Rita Rismala and Said Al Faraby, "Prediksi Tingkat Inflasi di Indonesia berbasis Jaringan Syaraf Tiruan dan Algoritma Genetika," in *Seminar Nasional Ilmu Komutasi dan Teknik Informatika*, 2013, p. 195.
- [7] Dr. G. Vadivu Navin, "Big Data Analytics for Gold Price Forecast Based on Decision Tree algorithm and Support Vector Regression (SVR)," *International Journal of Science and Research (IJSR)*, vol. Volume 4 , no. 3, March 2015, p. 2026, Maret 2015.
- [8] Robin Genuer, Jean-Michel Poggi, Christine Tuleau-Malot, and Nathalie Villa-Vialaneix, "Random Forest for Big Data," vol. arXiv:1511.08327, March 2017. [Online]. <https://arxiv.org/abs/1511.08327>
- [9] Ibrahim Yonis Omar, Romain Laborde, Ahmad Samer Wazan, Francois Barrere, Abdelmalek Benzekri, "G-Cloud on Openstack : Addressing access control and regulation requirements," in *Networks, Computers and Communications (ISNCC)*, Hammamet, Tunisia, 2015.
- [10] Ibrahim Abaker Targio Hashem; Ibrar Yaqoob; Nor Badrul Anuar; salimah Mokhtar; Abdullah Gani, Samee Ullah Khan, "The rise of "big data" on cloud Computing ," *Information System*, vol. 47, pp. 98-115, Juni 2014.

- [11] Budi Santosa, *Data Mining Teknik Pemanfaatan untuk Keperluan Bisnis*. Indonesia: Graha Ilmu, 2007.
- [12] (2016, November) DosenEkonomi.com Portal Studi Ekonomi. [Online]. <http://dosenekonomi.com/ilmu-ekonomi/cara-menghitung-laju-inflasi>
- [13] Soffa Zahara, "Interoperabilitas Tingkat Fungsionalitas Aplikasi Pada Migrasi Virtual Machine Di Lingkungan Cloud Computing," , 2017.
- [14] P.Sowjanya T. Madhuri, "Microsoft Azure v/s Amazon AWS Cloud Services: A Comparative Study," *International Journal of Innovative Research in Science Engineering and Technology*, vol. 5, no. 3, p. 3904, march 2016.
- [15] Alexis Wallskog Pappas and Universitet UMEA. (2014) Migration of Legacy Applications to the Cloud.
- [16] Ekaterina Zankova, "High Frequency Financial Time Series Prediction : machine learning approach," May 2016.
- [17] Ana Beatriz Galv.,o Michael P. Clements, "Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US output growth," in *International Symposium of Forecasting*, Santander, May 2007.
- [18] M. El-Dereny and N. I. Rashwan, "Solving Multicollinearity Problem Using Ridge Regression Models ," *Int. J. Contemp. Math. Sciences*, Vol. 6, 2011, no. 12, 585 - 600 , vol. 6 no. 12, pp. 585 - 600 , 2011.
- [19] (2010-2016) Scikit learn. [Online]. <http://scikit-learn.org>
- [20] Feedback Instruments Ltd., *Digital Pendulum: Control in a Matlab Environment*. Sussex, UK: Feedback Instruments Ltd., 2006.
- [21] Kazuo Tanaka and Michio Sugeno, "Stability analysis and design of fuzzy control," *Fuzzy Sets and Systems*, vol. 45, pp. 135-156, 1992.
- [22] Honghai F., Guoshun C., Cheng Y., Bingru Y., and Yumei C., "A SVM Regression Based Approach to Filling in Missing Values," in *Knowledge-Based Intelligent Information and Engineering Systems. KES 2005. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2005, pp. 581-587.
- [23] Preston J. Miller; Daniel M. Chin, "Using Monthly Data to Improve Quartely Model Forecast," *Spring* , vol. Vol. 20, No. 2, no. ISSN 0271-5287, July 1996.
- [24] Preston J. Miller and Daniel M. Chin, "Using Monthly Data to Improve Quartely Model Forecasts," *Quartely Review*, vol. 20. , no. Federal Reserve

Bank of Minneapolis, p. 16, 1996.

- [25] Michael P. Clements and Ana Beatric Galvao, "Macroeconomics Forecasting with Mixed Frequency Data : Forecasting US output growth," in *International Symposium of Forecasting*, Santander, 2007.

Halaman ini sengaja dikosongkan

LAMPIRAN

Lampiran 1. Data Surabaya bulan Januari 2012 s.d 2016

no	Tanggal	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
1	Jan-12	9180	7652	9641	23997	10501	60952	23261	41797	15565	8000	7809	5047	399	3000	7002	1498
2	Feb-12	9206	7703	9852	24000	10501	61641	23250	42145	15972	8000	7868	5013	400	3000	7004	1510
3	Mar-12	9207	7669	10733	25019	11114	62197	21805	41361	15019	8000	7930	5006	395	3000	6997	1505
4	Apr-12	9226	7615	10975	26993	11429	62833	20306	40833	13685	7990	7990	5030	384	3000	7011	1490
5	May-12	9094	7660	11600	27000	11452	62912	21981	40580	13649	8000	8000	5038	380	3000	7000	1485
6	Jun-12	9176	7725	12273	26973	11031	62847	22162	41107	14869	8000	8000	5035	376	3000	7004	1484
7	Jul-12	9261	7747	12174	26871	10735	64403	24326	42794	16849	8000	8000	5039	376	3000	7052	1504
8	Aug-12	9253	7740	11328	26400	10878	68755	25083	44065	14605	8000	8000	5037	376	3000	7000	1505
9	Sep-12	9240	7683	11183	26400	10705	68680	23135	42740	13986	8000	8000	5029	376	3000	7000	1500
10	Oct-12	9258	7715	11037	26400	9766	71900	24241	43958	13770	8000	8000	5131	376	3003	7000	1500
11	Nov-12	9285	7776	11359	26400	9279	75397	22324	42907	14261	8000	8000	5023	376	3100	7000	1500
12	Dec-12	9342	7808	11161	26400	8923	83290	22501	43671	15647	8000	8000	5119	376	3100	7000	1500
13	Jan-13	9497	7800	10985	26400	9060	81035	25939	45245	17830	8000	8000	5249	376	3100	7000	1513
14	Feb-13	9600	7800	11059	26400	9281	80421	26030	46654	18034	8000	8000	5320	376	3115	7000	1540

Lampiran 2. Data Surabaya bulan Januari 2012 s.d 2016 (Lanjutan)

no	Tanggal	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30	IHK
1	Jan-12	20019	13597	6609	8261	42271	13254	14971	2994	5370	7441	10557	6338	5488	9160	98.2
2	Feb-12	11652	14845	8436	7870	42307	13012	14849	2942	5537	7412	6550	5756	6177	9071	98.55
3	Mar-12	17937	27374	9670	8281	41813	12888	15597	2997	4883	7366	6007	5435	6528	9211	98.77
4	Apr-12	14047	28383	9199	9769	41833	13043	16001	2997	4328	7291	6300	4775	5249	9222	99.01
5	May-12	14504	11555	12054	10772	40617	13052	16252	3000	3885	7004	4601	4520	4851	9337	98.85
6	Jun-12	18513	14501	13061	14867	40980	13520	16827	3000	3725	6867	5206	4611	5197	9498	99.34
7	Jul-12	18252	13939	9068	15545	41187	13205	16748	3000	3723	6655	4825	4253	5271	9504	100.06
8	Aug-12	16055	16113	8674	15874	41603	13266	15990	3000	4061	7552	4966	5239	5672	9547	101.19
9	Sep-12	13141	14339	8143	16241	41897	12976	16143	3008	3901	7013	5410	5808	5896	9614	101.04
10	Oct-12	13793	14374	8230	13952	42506	11692	16019	3002	4710	6949	5046	7018	6918	9645	101.03
11	Nov-12	11357	11006	8955	15064	42600	11483	16137	3035	5489	7047	4777	8982	7261	9676	101.2
12	Dec-12	11022	11278	12975	15448	42639	11599	16662	3006	5978	7430	4871	9474	6849	9694	101.54
13	Jan-13	16875	22956	14368	21139	46232	11713	16919	3103	5158	8257	6212	9700	7010	9735	102.95
14	Feb-13	16142	23065	17386	29189	49607	12446	21921	3464	4316	8795	9100	9593	6893	9735	103.73

Lampiran 3 Data Surabaya bulan Januari 2012 s.d 2016 (Lanjutan)

no	Tanggal	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
15	Mar-13	9634	7800	11081	26400	9378	80226	25073	48403	15874	8000	8000	5334	376	3120	7000	1540
16	Apr-13	9626	7808	11109	26400	9285	79400	23145	45933	15071	8000	8000	5467	376	3120	7000	1540
17	May-13	9638	7920	11431	26400	9266	80071	23997	46110	16200	8000	8000	5550	376	3120	7000	1639
18	Jun-13	9637	7931	11310	26579	9583	80940	25065	47573	17477	8052	8052	5447	374	3122	7163	1700
19	Jul-13	9763	8014	11341	26170	9723	87000	30270	50184	19039	8416	8357	5639	370	3199	7408	1817
20	Aug-13	9603	7872	10997	23839	9830	87065	30098	50984	18008	8606	8254	5358	441	3213	7495	1741
21	Sep-13	9635	7914	10990	23540	10551	84747	30062	51307	16581	8699	8296	5173	442	3288	7260	1760
22	Oct-13	9868	7841	10780	23423	10564	84416	27653	47620	15221	8660	8300	5190	442	3360	6932	1769
23	Nov-13	9763	7844	10553	23573	11065	85207	26104	46425	14830	8620	8300	5160	437	3364	7058	1791
24	Dec-13	9744	7894	10610	24005	11266	87819	25272	48145	15499	8403	8161	4000	600	2600	7650	1676
25	Jan-14	9828	7908	10510	24964	11354	91142	27539	48552	17461	8275	8313	5200	410	3320	7172	1800
26	Feb-14	9947	7989	10685	25679	11663	91889	27225	49307	17734	8280	8380	5200	410	3360	7285	1800
27	Mar-14	10007	8086	10681	26300	12091	92732	24294	49597	14287	8988	9074	5200	410	3366	7434	1886
28	Apr-14	9872	8067	10503	26300	11683	93030	24528	49723	14493	9081	9217	5202	410	3364	7561	1900
29	May-14	9924	8021	10246	26300	11726	92071	26439	48613	16401	9083	9317	5158	410	3343	7364	1900
30	Jun-14	9912	8063	10106	26307	11839	94893	28698	50533	18092	9112	9273	5464	410	3410	7509	1900

Lampiran 4 Data Surabaya bulan Januari 2012 s.d 2016 (Lanjutan)

no	Tanggal	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30	IHK
15	Mar-13	15935	35874	35361	41497	52090	12929	22677	3285	4651	8551	9290	8023	6403	9758	104.73
16	Apr-13	17843	28433	34353	16697	46123	13010	23197	3300	5062	7577	7547	6467	6557	9773	104.34
17	May-13	19368	15658	25332	13791	47052	13398	23400	3300	5702	7894	6710	7071	6561	9810	103.6
18	Jun-13	23810	25743	22647	13811	48873	13981	22387	3323	6077	8134	7590	8443	6647	9931	104.41
19	Jul-13	23100	58555	41048	11311	50790	15655	19481	3667	6297	8821	10313	10550	7210	10124	108.17
20	Aug-13	23839	45144	42684	8752	55177	15620	16910	3537	5455	10137	11155	9950	6626	10625	109.17
21	Sep-13	17630	24580	23483	8323	52733	15193	16941	3402	4017	8883	6237	6884	5393	11403	108.35
22	Oct-13	29694	33716	20984	9187	48613	15623	16002	3242	4115	8845	4966	7039	6081	11424	108.56
23	Nov-13	19780	20283	24733	9583	47607	15888	16477	3326	3923	9335	4503	6827	6423	11671	108.47
24	Dec-13	26161	13806	25613	10435	45000	14065	14500	2900	3185	8452	7516	3863	7573	12148	109.18
25	Jan-14	26565	25132	19752	9335	49090	16028	16442	3396	5682	9477	8371	7776	7204	12238	110.47
26	Feb-14	23107	34414	14973	9620	49661	16403	16519	3557	6514	9675	7079	8275	7047	11995	110.72
27	Mar-14	17719	48168	13826	10884	49500	16489	16558	3400	5855	9945	3497	8100	7081	11484	110.97
28	Apr-14	11670	39183	14283	10537	50850	16971	16630	3537	4367	8757	3367	8087	6913	11493	111.16
29	May-14	9281	12323	14165	10129	51194	18491	16655	3300	4095	8626	6023	7755	6700	11584	111.35
30	Jun-14	9058	10003	18477	12363	51483	18354	16715	3350	4303	9493	6433	8120	7600	11952	111.76

Lampiran 5 Data Surabaya bulan Januari 2012 s.d 2016 (Lanjutan)

no	Tanggal	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
31	Jul-14	9956	8120	9828	26320	11904	94781	27684	52000	18371	9240	9160	5480	410	3420	7511	1900
32	Aug-14	9906	6500	9750	21944	11653	108387	30468	72581	16919	8500	8000	4097	600	2600	7605	2000
33	Sep-14	9907	8218	10116	26863	11357	94353	29970	52280	17607	9240	9160	5480	410	3330	7600	1900
34	Oct-14	9901	8220	9802	27323	11163	93106	25939	52400	16875	9240	9160	5286	410	3320	7634	1900
35	Nov-14	9889	8239	9930	27687	11303	93720	25271	52237	16267	8867	8367	4483	603	2593	7500	2000
36	Dec-14	9977	8483	9574	27600	10929	94435	25365	52839	18520	9646	9493	5495	398	3229	7624	1902
37	Jan-15	10139	8832	9511	27600	10953	94119	28694	53000	20520	9827	9667	5516	400	3230	7566	1955
38	Feb-15	10078	9044	9396	27471	10904	93857	29011	52307	19575	9840	9680	5534	400	3227	7569	1960
39	Mar-15	10154	8756	9612	27400	11017	93832	26886	52400	17308	9840	9707	5508	404	3219	7540	1960
40	Apr-15	10192	8599	10200	27107	11072	93990	26047	52280	16802	9793	9791	5617	401	3360	7512	1960
41	May-15	10415	8683	11114	27213	10913	94558	27768	52077	18763	9749	9747	5594	400	3602	7540	1960
42	Jun-15	10594	8631	11813	27223	11467	95825	29262	53010	20308	9724	9750	5789	403	3627	7607	1960
43	Jul-15	10482	8571	11348	27277	11073	99897	31144	54955	18494	9720	9750	5664	400	3660	8317	1960
44	Aug-15	10459	8625	10978	27600	10736	98242	33442	57987	20506	9720	9750	5706	400	3660	8331	1960
45	Sep-15	10582	8825	10969	27813	10572	96643	28325	54900	18884	9720	9750	5900	400	3750	8349	1960

Lampiran 6 Data Surabaya bulan Januari 2012 s.d 2016 (Lanjutan)

no	Tanggal	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30	IHK
31	Jul-14	9790	9990	18903	12358	52000	18129	17341	3680	4139	9606	6668	8177	7942	11748	112.23
32	Aug-14	8758	6806	14452	12048	45000	16597	15226	2900	4718	9073	4024	7435	6089	11765	112.79
33	Sep-14	11547	12757	13320	10308	52360	17203	16626	3480	5847	10087	5810	8483	7433	11950	113.25
34	Oct-14	15374	12739	13965	10543	52884	16916	16037	3559	8316	10129	5694	8360	7200	12206	113.8
35	Nov-14	41733	36900	14367	12900	45167	14917	16400	2897	5400	8350	5300	7283	8250	12219	115.24
36	Dec-14	65400	70145	15163	13486	58065	17165	17806	3886	5655	10148	6681	7884	13755	12500	117.81
37	Jan-15	39855	46554	14819	13006	58000	17048	18100	3803	5437	10373	6935	10335	9819	12514	118.29
38	Feb-15	15843	20911	14050	12864	57357	17257	18486	3794	5361	10046	6493	11243	8632	12686	117.79
39	Mar-15	14526	25329	20655	13923	57000	18016	18997	3800	5101	9716	6477	10574	7916	13002	118.21
40	Apr-15	13890	17847	23427	14191	56700	18100	19278	3785	4813	9617	7440	9560	7980	12883	118.69
41	May-15	21219	19884	25626	15732	56077	18100	19558	3826	4984	9310	8632	8729	7939	13075	119.15
42	Jun-15	23532	18673	23677	17430	56400	18793	20187	3800	5395	9495	7997	8457	8690	13247	119.79
43	Jul-15	27261	35394	18268	16955	57174	19310	20748	3800	5690	9627	6671	9157	9216	13308	120.25
44	Aug-15	24168	57768	13929	16897	57594	19539	20726	3800	5585	9571	6065	9821	9474	13713	120.83
45	Sep-15	20803	42367	14093	18097	56720	19707	20913	3753	5685	9817	5753	10181	9358	14324	121.14

Lampiran 7 Data Surabaya bulan Januari 2012 s.d 2016 (Lanjutan)

no	Tanggal	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16
46	Oct-15	10814	9279	10943	27929	10792	97029	27206	53148	17848	9720	9750	5874	400	3702	8419	1960
47	Nov-15	11429	9656	10899	28000	10784	98610	28643	53700	18382	9720	9750	5900	418	3744	8440	1960
48	Dec-15	11680	9800	11916	28000	10585	99116	30426	53329	21065	9720	9750	6041	402	3712	8437	1960
49	Jan-16	11520	9851	12112	27910	10338	104303	32668	57135	23074	9720	9750	6566	400	3700	8390	1995
50	Feb-16	11379	9946	12084	27600	10346	106445	32171	55897	21953	9720	9750	6797	400	3700	8420	2000
51	Mar-16	11185	9802	12085	27471	10132	105787	27984	55000	18919	9720	9750	7271	400	3700	8420	2000
52	Apr-16	11248	9076	12687	26313	11192	105590	28170	53433	18821	9979	9607	7293	729	3391	7903	2043
53	May-16	11241	9176	14729	26268	12280	108012	28714	53226	19297	10012	9526	7093	914	3240	7865	2077
54	Jun-16	11245	9565	15653	26689	11895	108332	30412	54372	20832	9952	9656	7069	853	2644	8014	2134
55	Jul-16	11328	9453	15744	26700	11275	109275	31729	59258	20119	9966	9742	7105	696	2954	8151	2215
56	Aug-16	12021	9300	14168	26139	11410	108387	29716	66419	20239	10306	9675	7192	648	3200	8648	2263
57	Sep-16	11897	9317	13580	25347	11890	109118	28973	65800	18967	10227	9620	7570	627	3200	8610	2260
58	Oct-16	12060	9452	13358	25552	12019	108462	29361	66581	18339	9900	9811	7527	620	3200	8506	2260
59	Nov-16	12245	9500	12738	25810	11775	110395	28747	69233	16823	9900	10079	7205	620	3200	8553	2249
60	Dec-16	12300	9500	12907	25800	11759	110411	29226	70710	20226	9900	10100	7097	620	3200	8500	2240

Lampiran 8 Data Surabaya bulan Januari 2012 s.d 2016 (Lanjutan)

no	Tanggal	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	X30	IHK
46	Oct-15	11726	16113	14645	19332	57000	19655	22103	3481	6097	10242	7913	13494	9758	13727	120.73
47	Nov-15	10490	12887	14073	20230	57000	19340	22627	3467	7947	10867	8180	12373	9767	13604	120.71
48	Dec-15	21071	26032	24432	22310	57000	19787	22894	3587	9297	11371	8219	14519	13842	13785	121.85
49	Jan-16	28023	28587	27171	25468	57000	19273	24977	3761	8903	17490	8616	13271	10639	13961	122.74
50	Feb-16	29714	17959	20816	25072	57000	19138	24848	3628	6903	12786	8752	12652	8866	13583	122.6
51	Mar-16	40361	38748	34916	29668	57000	19227	24777	3600	6671	10413	10887	10952	10139	13259	122.67
52	Apr-16	25150	25962	35247	32360	61223	19727	24610	4600	7148	9658	15483	10327	12340	13246	122.49
53	May-16	23773	21928	36916	33945	63258	19965	24906	4990	7837	10314	12288	11173	10084	13487	122.65
54	Jun-16	22745	22154	30339	33154	66317	19212	25797	5185	7199	12883	9222	13826	11477	13422	123.5
55	Jul-16	24645	31135	33581	32974	70887	19558	26368	4890	7290	15168	9316	13555	10935	13185	124.53
56	Aug-16	27200	36387	33632	31284	71871	20645	26497	4800	6465	15890	9452	10871	10439	13231	124.65
57	Sep-16	31380	29740	36433	32067	73333	18923	25793	4813	6047	16023	9967	9853	9873	13184	124.88
58	Oct-16	43374	28348	31316	32490	71355	18097	25148	4710	5903	15335	11258	9929	9877	13082	124.75
59	Nov-16	52573	46560	41073	34813	68900	17607	24387	4680	7473	12900	11447	8100	13087	13377	125.07
60	Dec-16	35832	51742	34142	34626	69774	17574	24877	4600	12703	12129	10935	10045	14432	13485	125.77

Lampiran 9 Keterangan Nama Variabel

No	Keterangan Nama variabel	No	Keterangan Nama variabel	No	Keterangan Nama variabel
x1	Beras Bengawan	x11	Susu Kental Manis Indomilk	x21	Ikan asin teri
x2	Beras IR 64	x12	Jagung pipilan kering	x22	Kacang hijau
x3	Gula Pasir Dalam Negri	x13	Garam bata	x23	Kacang tanah
x4	Minyak Bimoli 2 Liter	x14	Garam halus	x24	Ketela pohon
x5	Minyak Tanpa Merk	x15	Tepung segitiga biru	x25	Kol/kubis
x6	Daging Sapi Murni	x16	Indomie rasa kari ayam	x26	Kentang
x7	Daging Ayam Broiler	x17	Cabai biasa	x27	Tomat
x8	Daging Ayam Kampung	x18	Cabai rawit	x28	Wortel
x9	Telur Ayam Ras / Petelur	x19	Bawang merah	x29	Buncis
x10	Susu Kental Manis Merk Bendera	x20	Bawang putih	x30	Kurs dollar

Lampiran 10 . Tabel Perbandingan Waktu Training dan testing

			PC		EC2		Azure	
Method 0.3			timetrain	timetest	timetrain	timetest	timetrain	timetest
Ridge Regression	42	18	0.0108	0.0008	0.0038	0.0005	0.0022	0.011
linear Regression			0.002	0.0003	0.0013	0.0002	0.0012	0.0002
SVR-linear			0.0019	0.0006	0.001	0.0002	0.0011	0.0002
SVR-RBF			0.0017	0.0004	0.0009	0.0002	0.0009	0.0002
SVR-Poly			0.001	0.0004	0.0009	0.0002	0.0009	0.0002
RFR			0.0726	0.0024	0.0172	0.0012	0.0286	0.0015
Method 0.2			timetrain	timetest	timetrain	timetest	timetrain	timetest
Ridge Regression	48		0.0074	0.0008	0.0036	0.0004	0.028	0.0004
linear Regression			0.0021	0.0003	0.0013	0.0002	0.0017	0.0002
SVR-linear			0.0022	0.0004	0.0012	0.0002	0.0012	0.0002
SVR-RBF			0.0141	0.0005	0.0012	0.0002	0.0011	0.0002
SVR-Poly			0.0022	0.0004	0.0011	0.0002	0.0011	0.0002
RFR			0.0877	0.0026	0.0192	0.0013	0.0219	0.0015
Method 0.1			timetrain	timetest	timetrain	timetest	timetrain	timetest
Ridge Regression	54		0.005	0.0008	0.001	0.0004	0.0033	0.0004
linear Regression			0.0026	0.0003	0.0011	0.0001	0.0015	0.0002
SVR-linear			0.0043	0.0005	0.0023	0.0002	0.0021	0.0002
SVR-RBF			0.0038	0.0004	0.0019	0.0002	0.0035	0.0002
SVR-Poly			0.017	0.0004	0.0024	0.0002	0.0025	0.0002
RFR			0.0918	0.0036	0.0144	0.0011	0.0265	0.0013

Lampiran 11. Tabel Perbandingan Waktu Training dan testing (Lanjutan)

			PC		EC2		Azure	
Method	n		timetrain	timetest	timetrain	timetest	timetrain	timetest
Ridge Regression	1278		0.1845	0.0799	0.0782	0.021	0.0673	0.0198
linear Regression			0.0054	0.0019	0.0034	0.001	0.0033	0.0008
SVR-linear			3.5175	0.0357	2.1519	0.0124	2.1083	0.0131
SVR-RBF			0.2481	0.0279	0.1486	0.0115	0.1537	0.0143
SVR-Poly			0.5081	0.0235	0.3119	0.0144	0.3087	0.0173
RFR			0.2142	0.0042	0.1336	0.0028	0.1292	0.0031
Method			timetrain	timetest	timetrain	timetest	timetrain	timetest
Ridge Regression	1461		0.2499	0.0358	0.113	0.0161	0.1156	0.0157
linear Regression			0.0152	0.0016	0.0039	0.0008	0.0057	0.0008
SVR-linear			0.6083	0.0242	0.3508	0.0101	0.3629	0.0101
SVR-RBF			0.2746	0.019	0.1737	0.0084	0.1674	0.008
SVR-Poly			0.6744	0.0184	0.4295	0.0118	0.416	0.0114
RFR			0.3115	0.0041	0.2023	0.0031	0.1859	0.0032
Method			timetrain	timetest	timetrain	timetest	timetrain	timetest
Ridge Regression	1644		0.5779	0.0201	0.4239	0.0092	0.1353	0.0091
linear Regression			0.0085	0.001	0.0048	0.0005	0.0056	0.0006
SVR-linear			0.8104	0.0139	0.4818	0.0059	0.4692	0.0058
SVR-RBF			0.3578	0.0092	0.2221	0.0045	0.2079	0.0044
SVR-Poly			0.7929	0.0114	0.4925	0.0069	0.5384	0.0065
RFR			0.3446	0.0033	0.2205	0.0026	0.1944	0.0029

Lampiran 12. Perbandingan Waktu komputasi Grid Search dan Random Search di Cloud

	PC				Amazon EC2				Azure			
N = 42	Grid	R2	Random	R2	Grid	R2	Random	R2	Grid	R2	Random	R2
Ridge Regression	0.703	0.988	0.929	0.981	0.213	0.987	0.193	0.987	0.268	0.98	0.275	0.975
SVR-linear	15.972	0.985	1.866	0.988	10.093	0.987	1.142	0.986	12.42	0.987	1.35	0.977
SVR-RBF	11.5	0.985	1.469	0.988	7.779	0.986	0.882	0.986	8.361	0.985	0.994	0.982
SVR-Poly	20.08	0.985	1.878	0.988	10.929	0.987	1.198	0.987	14.133	0.987	1.283	0.977
RFR	156.023	0.984	10.037	0.983	111.721	0.99	7.757	0.99	126.658	0.982	7.941	0.979
N = 48	Grid	R2	Random	R2	Grid	R2	Random	R2	Grid	R2	Random	R2
Ridge Regression	0.783	0.985	0.494	0.97	0.193 13.64	0.94	0.12	0.922	0.324 13.177	0.985	0.188	0.922
SVR-linear	20.348	0.988	2.409	0.972		0.966	1.631	0.966		0.989	1.68	0.947
SVR-RBF	11.638	0.992	1.354	0.987	9.157	0.981	0.943	0.981	8.903	0.991	0.993	0.966
SVR-Poly	20.541	0.988	2.411	0.972	14.746	0.967	1.318	0.965	14.914	0.989	1.522	0.946
RFR	154.468	0.991	10.864	0.986	112.647	0.985	7.88	0.982	136.7	0.991	7.901	0.983
N = 54	Grid	R2	Random	R2	Grid	R2	Random	R2	Grid	R2	Random	R2
Ridge Regression	0.682	0.99	0.553	0.981	0.201 13.518	0.98	0.205	0.98	0.312 21.242	0.987	0.173	0.981
SVR-linear	27.601	0.992	3.084	0.989		0.988	1.626	0.988		0.989	1.955	0.988
SVR-RBF	12.418	0.993	1.593	0.992	8.905	0.989	0.99	0.989	9.426	0.988	1.058	0.988
SVR-Poly	27.344	0.992	2.845	0.989	14.957	0.988	1.421	0.988	20.517	0.989	1.715	0.988
RFR	167.027	0.99	11.632	0.992	116.082	0.987	7.356	0.984	134.263	0.985	8.281	0.987

Lampiran 13 Perbandingan Waktu komputasi Grid Search dan Random Search di Cloud (Lanjutan)

	PC				Amazon EC2				Azure			
N=1278	Grid	R2	Random	R2	Grid	R2	Random	R2	Grid	R2	Random	R2
Ridge Regression	30.341	0.999	28.856	0.999	9.662	0.999	9.66	0.999	8.768	0.999	8.564	0.999
SVR-linear	1947.97	0.997	186.124	0.997	1195.27	0.997	111.84	0.997	1274.89	0.997	200.851	0.997
SVR-RBF	594.054	1	56.38	1	361.663	1	36.46	1	373.487	1	44.863	1
SVR-Poly	2038.74	0.997	1018.55	0.999	1273.37	0.997	922.874	0.999	1318.72	0.997	1388.5	0.999
RFR	681.24	0.999	42.368	0.999	435.251	0.999	28.776	0.999	467.493	0.999	30.768	0.999
N = 1461	Grid	R2	Random	R2	Grid	R2	Random	R2	Grid	R2	Random	R2
Ridge Regression	32.822	0.999	26.65	0.999	12.486	0.999	9.485	0.999	11.641	0.999	6.526	0.999
SVR-linear	2309.57	0.997	293.646	0.997	1444.25	0.997	136.607	0.997	1588.77	0.997	185.076	0.997
SVR-RBF	700.065	1	74.519	1	439.184	1	52.622	1	452.608	1	48.721	1
SVR-Poly	2428.03	0.997	546.557	0.999	1505.81	0.999	958.86	1	1626.12	0.997	1399.89	0.999
RFR	737.21	0.999	46.828	0.999	483.927	0.999	34.339	0.999	503.689	0.999	30.111	0.999
N=1644	Grid	R2	Random	R2	Grid	R2	Random	R2	Grid	R2	Random	R2
Ridge Regression	38.734	0.999	51.714	0.999	15.776	0.999	16.354	0.999	14.15	0.999	8.782	0.999
SVR-linear	2833.641	0.997	328.863	0.997	1766.56	0.997	180.269	0.997	1875.75	0.997	236.309	0.997
SVR-RBF	836.538	1	87.787	1	521.763	1	69.21	1	526.433	1	51.652	1
SVR-Poly	2970.814	0.997	1672.866	1	1849.29	0.997	750.044	1	1949.71	0.997	670.331	0.999
RFR	798.492	0.999	60.936	0.999	521.748	0.999	34.865	0.999	530.346	0.999	35.313	0.999

Halaman ini sengaja dikosongkan

BIOGRAFI PENULIS



Intan Ari Budiastuti, anak pertama dari dua bersaudara. Lahir di Mojokerto, Jawa Timur. Dari kecil hingga menamatkan SMA berada di kota Mojokerto. Sekolah dasar di SDN Gedongan 3, sekolah menengah pertama di SLTPN 1 kota Mojokerto, dan sekolah menengah atas di SMUN 1 Sooko Mojokerto. Ia menyelesaikan Sarjana Teknik Informatika di Sekolah Tinggi Teknologi Telkom (STT Telkom) di Bandung pada tahun 2004.

Pengalaman kerja antara lain sebagai Staf programmer di perusahaan swasta Sisindosat di Jakarta dari 2004 sampai dengan 2006. Dan pada tahun 2006, menjadi Pegawai Negeri Sipil sebagai staf Data Elektronik, perangkat keras dan Jaringan sampai dengan sekarang di Dinas Komunikasi dan Informatika Pemerintah Daerah Kota Mojokerto. Tahun 2015 mengikuti program beasiswa pendidikan S2 dari Kementerian Komunikasi dan Informasi Republik Indonesia bidang Telematika CIO (Chief Information Officer) di Teknik Elektro ITS Surabaya. Dapat dihubungi melalui e-mail: asyf4h@gmail.com.

